
MASTERARBEIT

Frau
Bianca Liebscher (B.Sc.)

SEAT
Ein Programm zur Simulation und
Analyse experimenteller
Massenspektren.

Mittweida, 2014

MASTERARBEIT

Ein Programm zur Simulation und Analyse experimenteller Massenspektren.

Autorin:
Frau Bianca Liebscher (B.Sc.)

Studiengang:
Molekularbiologie/Bioinformatik

Seminargruppe:
MO11w1-B

Erstprüfer:
Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer:
Dr. Marius Kroll

Einreichung:
Mittweida, 28. Februar 2014

Verteidigung/Bewertung:
Mittweida, 2014

MASTER THESIS

SEAT A program to simulate and analyze experimental mass spectra.

author:

Ms. Bianca Liebscher (B.Sc.)

course of studies:

Molekularbiologie/Bioinformatik

tutorial group:

MO11w1-B

first examiner:

Prof. Dr. rer. nat. Dirk Labudde

second examiner:

Dr. Marius Kroll

submission:

Mittweida, February 28, 2014

defence/ evaluation:

Mittweida, 2014

Bibliografische Beschreibung:

Liebscher, Bianca:

SEAT - Ein Programm zur Simulation und Analyse experimenteller Massenspektren. -
2014 - 17, 95, 9 S.

Mittweida, Hochschule Mittweida, Fakultät Mathematik, Naturwissenschaften,
Informatik, Masterarbeit, 2014

Referat:

Ziel der Arbeit ist es Algorithmen für die Simulation von Profil- und Massenspektren zu entwickeln und zu implementieren. Es soll ein Programm entwickelt werden mit dem eine einfache Berechnung und Vergleich von simulierten und experimentellen Spektren möglich ist.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	V
Tabellenverzeichnis.....	VII
Formelverzeichnis	VIII
Abkürzungsverzeichnis	XI
1 Einleitung	1
2 Massenspektrometrie	2
2.1 Begriffsdefinitionen.....	2
2.2 Das Massenspektrometer	2
2.2.1 Ionenquellen und Ionisation	4
2.2.2 Elektrospray-Ionisation.....	6
2.2.3 Massenanalysatoren	9
2.3 Das Massenspektrum.....	22
2.3.1 Allgemeine Beschreibung.....	22
2.3.2 Signal-zu-Rausch-Verhältnis.....	23
2.3.3 Massenauflösung.....	23
2.3.4 Massengenauigkeit	25
2.3.5 Peak - Formen	25
2.3.6 Zusammenhang zwischen Peak-Abstand und Ladung.....	26
3 Simulation von Massenspektren.....	27
3.1 Elemente und ihre Isotope	27
3.1.1 Definition Isotop	27
3.1.2 Isotopenhäufigkeiten	27
3.2 Isotopenmuster im Massenspektrum.....	28
3.3 Berechnung von Molekül- und Ionenmassen.....	30
3.3.1 Nominelle Masse.....	31
3.3.2 Relative Masse	31
3.3.3 Exakte Masse	32
3.3.4 Masse-zu-Ladungsverhältnis	33
3.3.5 Der Massendefekt.....	33
3.4 Algorithmen zur Isotopenmusterberechnung.....	35
3.4.1 Binomiale und Polynomiale Methode	35
3.4.2 Multinomiale Methode	36
3.4.3 Fourier-Transformation basierte Berechnungen	37

3.4.4	Zweidimensionale Fourier-Transformation	41
4	Datensatz und Dateiformate	44
4.1	Spektren Datensatz	44
4.1.1	Massenspektrometer und Methode	44
4.1.2	Analyt	44
4.2	Datenformate	45
4.2.1	Aufbau der TXT-Datei	45
4.2.2	Aufbau der CSV-Datei	46
4.2.3	Aufbau der mzXML-Datei	47
4.3	Software	47
5	SEAT	49
5.1	Basisfunktionen von SEAT	49
5.1.1	Import von TXT, CSV und MZXML Dateien	49
5.1.2	Approximierte Auflösung für Spektren und Scans	52
5.1.3	Grafische Darstellung	55
5.2	Berechnung simulierter Spektren	56
5.2.1	Multinomiale Methode	56
5.2.2	2D Fourier Transformation - <i>Deuterium</i>	64
5.3	Vergleich von Spektren	67
5.3.1	Peak-Modell	67
5.3.2	Begriffe für die Klassifikation von Peaks	68
5.3.3	Distanz und Ähnlichkeit zwischen zwei Peaks	69
5.3.4	Distanzmatrix und Klassifikation der Peak-Paare	71
5.3.5	Ähnlichkeit zwischen Spektren	72
6	Beispielberechnungen mit SEAT	73
6.1	Simulation mit der multinomialen Methode	73
6.1.1	SAIB Spektren	73
6.1.2	Laufzeitmessungen	75
6.2	Simulationen mit der 2D-FT Methode	78
6.2.1	SAIB Spektren	78
6.2.2	Laufzeitmessungen	80
6.3	Vergleich von Spektren	82
6.3.1	Anwendung auf gemessene und simulierte Spektren	82
6.3.2	Analyse der Jeffrey-Matusita Distanz	84
7	Auswertung und Diskussion	86
7.1	Simulation der Spektren	86

7.1.1	Vergleich der Eingabe- und Ausgabe-Parameter	86
7.1.2	Laufzeit, Effizienz und Genauigkeit	87
7.1.3	Parameterwahl.....	91
7.2	Vergleich von Spektren	93
7.2.1	Parameterwahl.....	93
8	Zusammenfassung und Ausblick.....	94
Literaturverzeichnis		XII
Anhang.....		XVII
Selbstständigkeitserklärung.....		XXVIII

Abbildungsverzeichnis

Abbildung 2-1: Schematischer Grundaufbau eines Massenspektrometers	3
Abbildung 2-2: Einteilung einiger Ionisationsverfahren nach Einsatzgebiet und Analyt. 6	
Abbildung 2-3: Schematische Darstellung der Elektrospray-Ionisation	8
Abbildung 2-4: Beispiele für die Architektur einer ICR-Zelle	13
Abbildung 2-5: Kreisbewegung einer Ions in einem Magnetfeld.....	15
Abbildung 2-6: Cyclotron-Frequenz und -Radius in Abhängigkeit vom m/z - Wert	15
Abbildung 2-7: Schema der Trajektorie von Ionen bei Anregung und Detektion	17
Abbildung 2-8: Ionen-Bewegungen in einer ICR-Zelle	19
Abbildung 2-9: Umwandlung des Transienten von der Zeit- in die Frequenzdomäne ..	20
Abbildung 2-10: Auswirkung einer kurzen Messzeit auf die Frequenzdomäne	21
Abbildung 2-11: Schema des Bruker Daltonics Solarix FT-ICR-MS	21
Abbildung 2-12: schematische Darstellung eines Massenspektrums.....	22
Abbildung 2-13: Massenauflösung mit der 10% - Tal Definition und FWHM Definition	24
Abbildung 2-14: Peak – Formen nach der Gauss- und Lorentz – Funktion.....	26
Abbildung 3-1: Isotopenmuster von Xenon	28
Abbildung 3-2: Isotopenmuster von Chloroform und Kaliumbromid	30
Abbildung 3-3: Massendefekt pro Nukleon aller Isotope	34
Abbildung 3-4: Komprimierung des Massenspektrums	41
Abbildung 3-5: Schematische Darstellung der 2D – FT Methode.....	43
Abbildung 4-1: Herstellung von SAIB	45
Abbildung 4-2: MZXML Format.....	47
Abbildung 5-1: Datenmodell	50
Abbildung 5-2: Import von TXT und CSV Datei	51
Abbildung 5-3: Import von MZXML und CSV	52
Abbildung 5-4: Ergebnis des Curve-Fitting eines experimentellen Massenspektrums	54
Abbildung 5-5: Funktionen für die Approximation der Auflösung von Spektren.....	54
Abbildung 5-6: Darstellung des nicht fragmentierten Spektrums als Profilspektrum.....	55
Abbildung 5-7: Schema der Berechnungsschritte der multinomialen Methode	64
Abbildung 5-8: 1D Ausgabe von Deuterium.....	66
Abbildung 5-9: Peak-Modell und Distanz.....	68
Abbildung 5-10: Klassifikation der Peaks.....	69
Abbildung 6-1: Nicht fragmentiertes Spektrum und multinomial simuliertes Spektrum im Vergleich	73

Abbildung 6-2: Anzahl berechneter Peaks in Abhängigkeit von Summenformel und Schwellwert.....	75
Abbildung 6-4: Laufzeit der multinomialen Methode in Abhängigkeit von Isotopen- und Atomanzahl.....	76
Abbildung 6-5: Rechenzeit der multinomialen Methode von fünf Summenformeltypen	77
Abbildung 6-6: Nicht fragmentiertes Spektrum & 2D-FT simuliertes Spektrum im Vergleich.....	78
Abbildung 6-7: Anzahl berechneter Peaks in Abhängigkeit von Summenformel und Auflösung.....	80
Abbildung 6-8: Laufzeit der 2D-FT Methode in Abhängigkeit von Isotopen- und Atomanzahl.....	81
Abbildung 6-9: Rechenzeit der 2D – FT Methode für verschiedene Summenformeln ..	82
Abbildung 6-10: Beispiel der Klassifikation eines Isotopenpeaks.....	83
Abbildung 6-11: Jeffrey Matusita Distanz Analyse.....	85
Abbildung 7-1: Vergleich eines mit der multinomialen und 2D-FT Methode berechneten Peaks.....	89
Abbildung 7-2: Vergleich 2D-FT Spektrum mit multinomial simuliertem Spektrum von SAIB.....	90
Abbildung 7-3: Vergleich eines berechneten und gemessenen Isotopenmusters	91
Abbildung 7-4: Ein mit 2D-FT simulierter Isotopenpeak verschiedener Auflösungen ...	92
Abbildung A-1: Anzahl berechneter Peaks pro Summenformel mit der multinomialen Methode des fragmentierten Spektrums.....	XXV
Abbildung A-2: Anzahl berechneter Peaks pro Summenformel mit der 2D-FT Methode des fragmentierten Spektrums	XXVI

Tabellenverzeichnis

Tabelle 2-1: Verschiedene durch ESI erzeugte Ionen.....	9
Tabelle 2-2: Zusammenfassung der Eigenschaften von Massenanalysatoren.....	10
Tabelle 3-1: Beispiel der unterschiedlichen Massendefinitionen	32
Tabelle 4-1: Aufbau der TXT-Dateien.....	45
Tabelle 4-2: Aufbau der CSV-Datei	46
Tabelle 5-1: Partitionen und deren Permutationen am Beispiel von 03	57
Tabelle 6-1: Berechnung der SAIB Spektren mit der multinomialen Methode.....	74
Tabelle 6-2: Ergebnisse der Berechnung der SAIB Spektren mit der 2D-FT Methode.....	79
Tabelle 6-3: Ergebnisse des Spektren-Vergleichs am Beispiel von SAIB	84
Tabelle 7-1: Vergleich der Eingabeparameter der multinomialen und 2D-FT Methode.....	86
Tabelle 7-2: Vergleich Ausgabeparameter der multinomialen und 2D-FT- Methode	87
Tabelle 7-3: Zusammenfassung und Parameterwahl für die Simulation von Spektren.....	92
Tabelle A-1: Isotopenzusammensetzung sowie Massen der Elemente	XVIII
Tabelle A-2: Laufzeiten der multinomialen und 2D-FT Methode für verschiedene Summenformeln.....	XXVII

Formelverzeichnis

Formel 2.1: Lorentz-Kraft.....	14
Formel 2.2: Zentrifugalkraft	14
Formel 2.3: Gleichgewicht zwischen Lorentz-Kraft und Zentrifugalkraft.....	14
Formel 2.4: Ionen Cyclotron Radius	14
Formel 2.5: Cyclotron-Winkelfrequenz.....	14
Formel 2.6: ungestörte Cyclotron-Frequenz	14
Formel 2.7: Cyclotron-Frequenz eines einfach geladenen Ions mit m/z 50, $B=9,4$ T ...	15
Formel 2.8: Elektrisches Feld für die Anregung der Ionen	16
Formel 2.9: Energieabsorption der Ionen durch Einwirkung eines elektrischen Feldes	16
Formel 2.10: Radius der Ionen nach einer Anregung	16
Formel 2.11: Radiale Kraft erzeugt durch quadrupolares Feld.....	18
Formel 2.12: Auf Ionen wirkende Gesamtkraft.....	18
Formel 2.13: Auf Ionen wirkende Gesamtkraft II.....	18
Formel 2.14: reduzierte Cyclotron-Frequenz	18
Formel 2.15: Magnetron-Frequenz	18
Formel 2.16: ungestörte Cyclotron-Frequenz	18
Formel 2.17: Massenauflösung R	23
Formel 2.18: Auflösung nach der 10%-Tal Regel	23
Formel 2.19: Auflösung nach der FWHM-Definition.....	24
Formel 2.20: Massengenauigkeit.....	25
Formel 2.21: Gauß-Funktion	25
Formel 2.22: Standardabweichung.....	25
Formel 2.23: Lorentz-Funktion	26
Formel 3.1: Berechnung der nominellen Masse eines Moleküls	31
Formel 3.2: Berechnung der relativen Masse eines Elements	31
Formel 3.3: Berechnung der genauen Isotopenmasse eines Moleküls	32
Formel 3.4: Berechnung der exakten Masse eines Ions	33
Formel 3.5: Berechnung der m/z Verhältnisses eines Ions	33
Formel 3.6: Massendefekt in der Massenspektrometrie	34
Formel 3.7: Binomiale Gleichung.....	35
Formel 3.8: Binomiale Methode am Beispiel von Wasserstoff	35
Formel 3.9: Polynomiale Darstellung einer Isotopenverteilung	36
Formel 3.10: Darstellung eines polynomialen Terms durch multinomiale Verteilung....	36

Formel 3.11: Vereinfachte Berechnung der Häufigkeit	36
Formel 3.12: Beispiel für die vereinfachte Berechnung der Häufigkeit	37
Formel 3.13: Einfügen eines Identifikators in die polynomiale Gleichung	38
Formel 3.14: Anwendung der multinomialen Methode	38
Formel 3.15: Anwendung der multinomialen Methode II	38
Formel 3.16: Gleichung der Isotopenverteilung in der μ -Domäne	38
Formel 3.17: Einbeziehen einer Peak-Form in die Fourier-Transformation	39
Formel 3.18: Heterodyning für das Verschieben der Masse nahe null	39
Formel 3.19: Massenbereich für die Fourier-Transformation	39
Formel 3.20: Komprimierung des Massenbereichs	40
Formel 3.21: Zweidimensionale Fourier-Transformation in μ Domäne	41
Formel 3.22: Zweidimensionale Fourier-Transformation in μ Domäne II	41
Formel 3.23: Zweidimensionale Fourier-Transformation in Massendomäne	42
Formel 5.1: Intervall Berechnung für Scan-Alignment	51
Formel 5.2: Auflösungs-Funktion LINEAR	53
Formel 5.3: Auflösungs-Funktion EXPONENTIAL 1	53
Formel 5.4: Auflösungs-Funktion EXPONENTIAL 2	53
Formel 5.5: Partition	56
Formel 5.6: Berechnung der Permutationen	60
Formel 5.7: Erweiterung der multinomialen Methode	61
Formel 5.8: Logarithmengesetze und die erweiterte multinomiale Methode	61
Formel 5.9: Berechnung der logarithmierten Fakultät	61
Formel 5.10: Kommandozeilenparameter von Deuterium	65
Formel 5.11: Befehl für die 2D-FT Methode	66
Formel 5.12: Beispiel für eine Simulation nach der 2D-FT Methode	66
Formel 5.13: Peak-Modell für Spektrenvergleich	67
Formel 5.14: Jeffrey – Matusita Distanz	70
Formel 5.15: Peak – Form für Jeffrey-Matusita Distanz	70
Formel 5.16: Jeffrey – Matusita Distanz mit Gauß - Verteilung	70
Formel 5.17: Bhattacharyya Distanz	70
Formel 5.18: Distanz aus Basis des ppm-Wertes	71
Formel 5.19: Distanzmatrix	71
Formel 5.20: Klassifikation-Funktion der Peak-Paare	72
Formel 5.21: Distanz zwischen zwei Spektren	72
Formel 5.22: Verhältnis von gepaarten und nicht gepaarten Peaks in den Spektren R und Q	72

Formel 6.1: Fünf Summenformeltypen für Laufzeitmessungen.....	76
Formel 7.1: Laufzeit im Worst-Case bei der Multinomialen Methode	87

Abkürzungsverzeichnis

APCI	Atmosphärendruck Chemische Ionisation
APPI	Atmosphärendruck Photoionisation
CI	Chemische Ionisation
CRM	Charged Residue Model
D	Deletion
Da	Dalton (Einheit)
EI	Elektronenstoßionisation
ESI	Elektrospray Ionisation
eV	Elektronenvolt (Einheit)
FAB	Fast Atom Bombardement
FD	Felddesorption
FI	Feldionisation
FT-ICR-MS	Fourier-Transform Ion-Cyclotron-Resonance Massenspektrometrie
I	Insertion
IEM	Ion Evaporation Model
M	Match/ Übereinstimmung
MM	Mismatch/ keine Übereinstimmung
MS	Massenspektrometrie
m/z	Masse-zu-Ladungs-Verhältnis
SAIB	Saccharoseacetatisobutyrat
u	unit (Einheit)

1 Einleitung

Seit der Entdeckung des Elektrons durch J. J. Thomson im Jahr 1897 [1] ist es möglich Massen von Elementen zu messen. Einige Jahre später gelang es F. W. Aston und A. J. Dempster die Isotope ^{20}Ne und ^{22}Ne nachzuweisen [2]. Diese frühen Experimente wurden mit Massenspektrographen, welche eine Fotoplatte zur Detektion benutzten, durchgeführt. Heute existieren viele verschiedene Massenspektrometer für die unterschiedlichsten Anwendungsgebiete in Medizin, Chemie, Biologie etc. Man kann mit ihnen die Zusammensetzungen verschiedener Analyten aufklären. Oft wird die Massenspektrometrie in Zusammenhang mit den „-omik – Wissenschaften“ genannt. So auch in der Petroleomik, die sich mit der Analyse von Erdöl und dessen Komponenten beschäftigt. Aufgrund der hochauflösenden *Fourier-Transform Ion Cyclotron Resonance* Massenspektrometrie ist es möglich hochkomplexe organische Gemische wie Erdöl zu analysieren und Informationen über die elementaren Zusammensetzungen von Stoffgruppen und einzelnen Molekülen zu gewinnen. Kennt man die genaue Zusammensetzung des Erdöls und anderer komplexer Gemische ist es möglich, die Eigenschaften und das Verhalten vorherzusagen. [3]

Aufgrund des immer steigenden Rohölbedarfs ist es wichtig andere Brennstoffquellen, aus zum Beispiel nachwachsenden Rohstoffen, herstellen zu können. Als eine mögliche Alternative sind die Pyrolyseöle zu nennen. Ihre Zusammensetzung ist ebenso komplex wie die des Erdöls jedoch können die Anzahl und Häufigkeit der Komponenten je nach Herstellung und verwendeten Rohstoffen sehr stark schwanken. Ebenso ist es wichtig, dass alle Komponenten bei der massenspektrometrischen Analyse intakt bleiben und nicht zerfallen.

An der technischen Universität Freiberg wird diese Problematik aufgegriffen und Pyrolyseöle mit Hilfe der *Fourier-Transform Ion Cyclotron Resonance* Massenspektrometrie untersucht.

Das Ziel dieser Arbeit liegt in der Implementierung verschiedener Algorithmen zur Simulation und dem Vergleich von Massenspektren. Des Weiteren soll ein Tool entwickelt werden, mit dem es möglich ist, Massenspektren grafisch darzustellen, mit ihren simulierten Varianten zu vergleichen und eine Aussage über die Qualität dieser zu erhalten.

2 Massenspektrometrie

Dieses Kapitel gibt eine kurze Übersicht über verschiedenen Methoden der Massenspektrometrie und beinhaltet wichtige Begriffsdefinitionen. Im letzten Abschnitt dieses Kapitels wird die Funktionsweise der FT-ICR-MS erläutert.

2.1 Begriffsdefinitionen

Als *Massenspektrometrie* bezeichnet man die Untersuchung an anorganischen und organischen Substanzen durch Erzeugung gasförmiger fragmentierter oder nicht fragmentierter Ionen. Diese Ionen werden durch ein spezifisches Masse-zu-Ladungs-Verhältnis, abgekürzt m/z , und ihrer Häufigkeit charakterisiert. [4, 5]

Das m/z - *Verhältnis* ist der dimensionslose Quotient aus der Massenzahl eines Ions und dessen Anzahl an Ladungen. Zum Beispiel besitzt ein zweifach positiv geladenes Benzolmolekül $C_6H_6^{2+}$ einen m/z - Wert von $(6 \cdot 12 + 6 \cdot 1)/2 = 39$. Bei $z = 1$ entspricht m/z direkt der Summe der Atommassen, aus denen das Ion zusammengesetzt ist. [6]

2.2 Das Massenspektrometer

Jedes Massenspektrometer ist aus den gleichen Grundbauteilen aufgebaut. Ein Probeneinlass nimmt die zu analysierenden Substanzen auf. Diese können entweder fest, flüssig oder gasförmig sein. Die Ionenquelle erzeugt die für die Messung benötigten Ionen. Je nach Aggregatzustand des Analyts stehen unterschiedliche Ionisationsmethoden zur Verfügung. Durch elektrische oder magnetische Felder im Massenanalysator werden die Ionen nach ihren m/z - Werten aufgetrennt und schließlich deren Häufigkeit im Detektor gemessen. Eine Ausnahme bilden die Time-of-Flight-Analysatoren, wo die Ionen in einer feldfreien Driftstrecke separiert werden. Für die Massenanalyse wird ein Hochvakuum benötigt, um mögliche Kollisionen mit in der Atmosphäre befindlichen Molekülen oder Atomen zu verhindern. Je nach Methode kann auch für die Ionisation ein Vakuum benötigt werden. Der allgemeine Aufbau eines Massenspektrometers ist in Abbildung 2-1 dargestellt. Als Ausgabe erhält man ein Massenspektrum auf dessen Abszisse das m/z - Verhältnis und auf der Ordinate die

Häufigkeit der Ionen eines bestimmten m/z - Wertes, der als Intensität bezeichnet wird, aufgetragen ist. [7, 5]

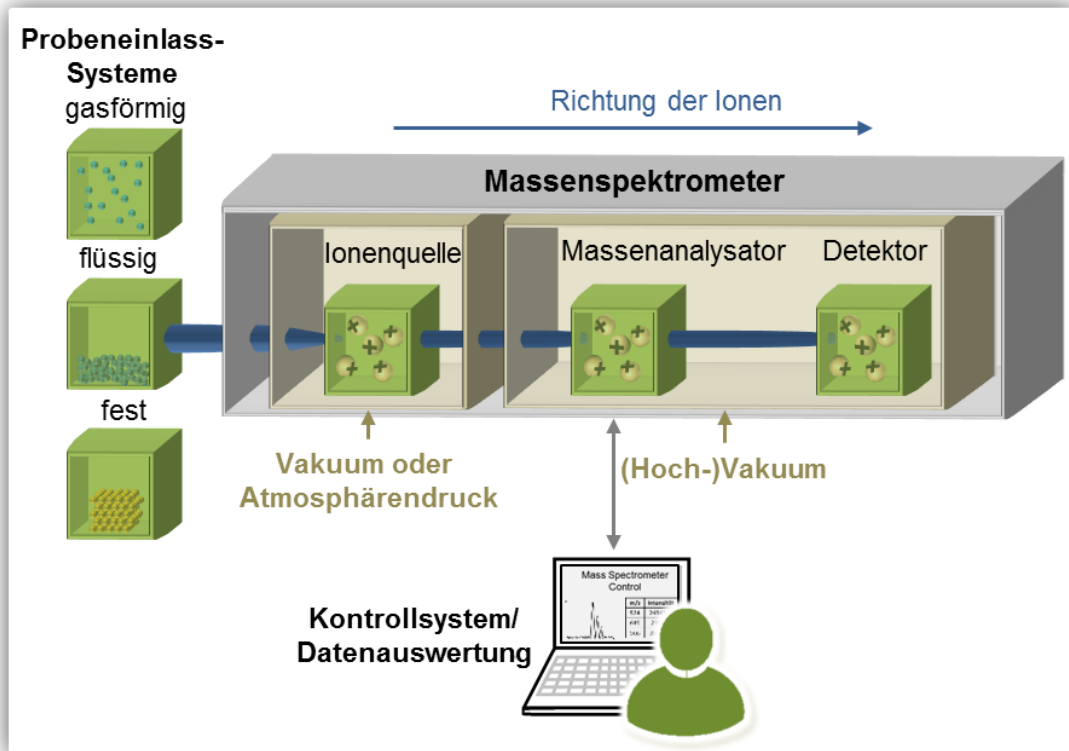


Abbildung 2-1: Schematischer Grundaufbau eines Massenspektrometers

Ein Massenspektrometer (grau) besteht aus drei Hauptkomponenten: Der Ionenquelle, dem Massenanalysator und dem Detektor. Die Massentrennung und Detektion erfolgt unter Vakuum und je nach Ionisierungsart auch die Erzeugung messbarer Ionen (braune Kästen). Für die Zuführung des Analyts wird ein Probeneinlasssystem benötigt. Je nach Ionisierungsart können Analyten aller Aggregatzustände untersucht werden. Ein Computersystem welches die Komponenten des Massenspektrometers überwacht und steuert zeichnet die Messwerte auf. In dieser stark vereinfachten Darstellung sind beispielhaft einfach positiv geladene Ionen dargestellt. (erstellt nach [5, 8])

2.2.1 Ionenquellen und Ionisation

Die Ionisation von Molekülen kann durch verschiedene physikalische und chemische Prozesse durchgeführt werden. Sie kann thermisch, durch elektrische Felder, durch Beschuss von energiereichen Elektronen, (atomare) Ionen oder Photonen, energiereichen neutralen Atomen, elektronisch angeregten Elektronen, Cluster-Ionen oder elektrostatisch aufgeladene Mikrotropfen erfolgen [5]. Bei einigen Methoden wirken hohe Energien, was Fragmentierungen des Analyten zur Folge hat. Je nach Aggregatzustand, Polarität und Molmasse des Analyten stehen unterschiedliche Verfahren zu Verfügung. Hier soll ein kurzer Überblick über einige Methoden gegeben werden:

Bei flüchtigen, thermisch stabilen Stoffen werden meist Elektronenstoßionisation (EI), Chemische Ionisation (CI) oder die Feldionisation (FI) verwendet. [5, 9]

Bei der *EI* werden für die Ionisation energiereiche Elektronen auf gasförmige, neutrale Analyt-Teilchen geschossen. Das geschieht unter Niederdruck im Bereich von 10^{-4} Pa . Genutzt wird diese Methode bei nieder- bis mittelpolaren, nicht ionischen Verbindungen. Die *CI* ist ähnlich der *EI*. Jedoch wird hier als Zwischenschritt ein Reaktandgas, wie zum Beispiel Methan, eingesetzt. Dieses wird durch Elektronenbeschuss ionisiert. Die entstandenen Reaktandgas-Ionen reagieren nun durch Molekül-Ion-Reaktionen mit dem Analyt in der Gasphase und übertragen Elektronen, Protonen oder andere geladene Teilchen. Werden gasförmige Moleküle einem starken elektrischen Feld ausgesetzt, können sich Elektronen lösen und so durch Autoionisation Ionen erzeugen. Diese Methode ist auch als *FI* bekannt und wird bei neutralen oder ionischen Analyten eingesetzt. [5, 9]

Will man Flüssigkeiten mit dem Massenspektrometer analysieren, müssen sie durch Zerstäubung in die Ionenquelle eingebracht werden. Die Ionisation wird dabei unter Atmosphärendruck durchgeführt. Durch entsprechende Vakuumpumpen werden die Ionen in den Massenanalysator weitergeleitet. Dazu gehören Elektrosprayionisation (ESI), Atmosphärendruck chemische Ionisation (APCI) und Atmosphärendruck-Photoionisation (APPI). Bei der *ESI* entstehen kleine geladenen Tropfen in einem elektrischen Feld. Die Verdampfung von Molekülen im Tropfen führen zur Reduktion der Tropfengröße und setzen am Ende die geladenen, desolvatisierten Analyt-Ionen frei. *ESI* wird meist bei großen Biomolekülen, Nucleinsäuren oder anderen Stoffen mit polaren Gruppen angewandt. Die *APCI* ist eine Variante der *CI*, bei der Ionen unter Atmosphärendruck gebildet werden. Diese Methode ist schonender und bietet Vorteile gegenüber der *ESI*, da mit dieser Methode auch niedere polare Stoffe untersucht werden

können. Die *APPI* ersetzt die bei der *APCI* eingesetzte Ionisationsquelle durch eine UV-Lichtquelle, die eine Ionisation mit Hilfe von Photonen auslöst. Oft werden *ESI*, *APPI* und *APCI* in einem Gerät gekoppelt sodass man zwischen den Methoden wählen kann. [5, 9]

Für die Analyse von nicht flüchtigen Feststoffen stehen die Matrix-unterstützte Laserdesorption/Ionisation (*MALDI*), die Felddesorption (*FD*), das Fast Atom Bombardment (*FAB*) oder die sekundäre Ionen Massenspektrometrie (*SIMS*) zur Verfügung. Das Grundprinzip dieser Methoden besteht aus der Benutzung eines festen Stoffes oder einer viskosen Flüssigkeit, die als Matrix, in der der Analyt eingebracht wird, eingesetzt wird. Werden Partikel mit sehr hohen kinetischen Energien auf eine Oberfläche geschossen, können Neutralteilchen oder Ionen herausgelöst werden. Das *FAB* zum Beispiel nutzt eine solche viskose Flüssigkeit als Matrix, die mit einem Strahl aus energiereichen Neutralteilchen beschossen wird. Bei der *SIMS* wird dafür ein Ionenstrahl verwendet. Bei der *FD* handelt es sich um eine Weiterentwicklung der *FI*, wobei die Ionisation nicht mehr in der Gasphase, sondern auf der Oberfläche der Feldanode, auch Feldemitter genannt, durchgeführt wird. Sie wird für neutrale oder ionische Analyten verwendet. *MALDI* nutzt eine Matrix, die aus kleinen organischen Molekülen besteht. In diese wird der Analyt gelöst und getrocknet, sodass eine feste kristalline Struktur entsteht. Anschließend wird mit einem Laser, mit einer bestimmten Wellenlänge auf die Matrix gezielt, sodass der Analyt aus der Matrix desorbiert und ionisiert wird. Dies findet unter Vakuum statt. *MALDI* wird oft für die Analyse großer Moleküle, wie Proteine, Oligonukleotide oder Polymere eingesetzt. [5, 9]

Einige dieser Ionisationsmethoden werden als „sanft“ oder „weich“ bezeichnet¹, da die Moleküle bei der Ionisation intakt bleiben und nicht fragmentieren. Mit diesen Methoden lassen zum Beispiel große (Bio-) Moleküle sehr gut analysieren. Im Gegensatz dazu fragmentieren Analyt-Moleküle bei „harten“ Ionisationsmethoden, wenn sie ionisiert werden. In Abbildung 2-2 sind die vorgestellten Ionisationsverfahren hinsichtlich ihrer „Härte“ und „Sanftheit“ und Einsatzgebiete eingeteilt. [5, 10]

¹ Engl.: soft ionization methods

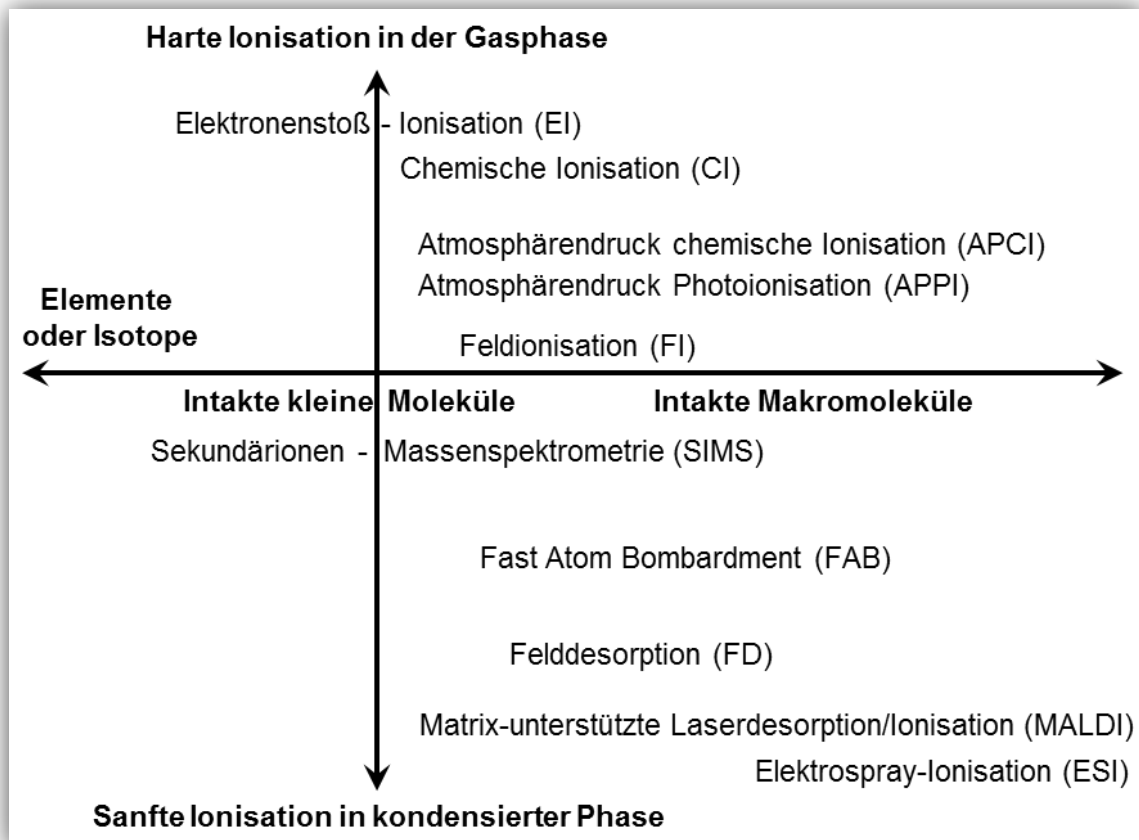


Abbildung 2-2: Einteilung einiger Ionisationsverfahren nach Einsatzgebiet und Analyt

Ionisationsverfahren können hinsichtlich ihrer „Härte“ oder „Sanftheit“ eingeteilt werden. Welche Methode ausgewählt wird, hängt von der Komplexität des Analyts und dessen Aggregatzustand ab. Diese Übersicht fasst die Verwendung der vorgestellten Ionisationsmethoden zusammen. (Abbildung erstellt nach [5], S. 8)

Im folgenden Abschnitt wird die sanfte Ionisationsmethode ESI näher erläutert.

2.2.2 Elektrospray-Ionisation

Eine der ersten Forschungen zu Elektrospray-Ionisation (ESI)² geht auf Malcolm Dole³ in den 70er Jahren zurück. Als Motivation diente die nicht vorhandene Möglichkeit, intakte Makromoleküle durch Verdampfung zu erhalten. Die Idee wurde von dem Chemiker John B. Fenn⁴ weiterentwickelt. Im Jahr 2002 gewann Fenn zusammen mit Koichi Tanaka den Nobelpreis für Chemie für die Entwicklung einer sanften Ionisationsmethode zur massenspektrometrischen Analyse biologischer Makromoleküle. [11, 12]

Bei der ESI handelt es sich um eine sanfte Ionisationsmethode die unter Atmosphärendruck durchgeführt wird. Im Gegensatz zu harten Ionisationsverfahren,

² In manchen älteren Literaturquellen wird die Elektrospray-Ionisation auch mit ES abgekürzt

³ Chemiker, * 04. März 1903, † 29. November 1990

⁴ Chemiker, * 15. Juni 1917, † 10. Dezember 2010

wie FAB oder EI, wird die Ionisationsenergie nicht direkt auf die zu ionisierenden Moleküle übertragen, sondern von „Ionisationspartnern“ absorbiert. Durch diese schonende Behandlung eignet sich ESI vor allem für hochmolekulare Stoffe wie Proteine und andere Biomoleküle, aber auch für kleine polare Moleküle, ionische Metallkomplexe oder lösliche anorganische Stoffe [5, 13]

Der Prozess der ESI kann in drei Schritten zusammengefasst werden: Zuerst wird ein elektrisch geladenes Spray erzeugt. Im zweiten Schritt wird die Größe der Tropfen sehr stark verringert um letztendlich freie gasförmige Ionen zu erhalten. Eine schematische Darstellung dieses Prozesses ist in Abbildung 2-3 abgebildet. Um einen polaren oder ionischen, nicht flüchtigen Analyten für eine ESI zugänglich zu machen, muss er mit einem flüchtigen Lösungsmittel vermischt werden. Je nach physikalisch-chemischen Eigenschaften des Analyten werden unterschiedliche Gemische benötigt, die zum Beispiel aus Wasser, Diethylether, Tetrahydrofuran, Dichlormethan, Methanol, Trichlormethan, Isopropanol oder Acetonitril etc. bestehen können. Die Konzentration des Analyten in der fertigen Lösung sollte bei 10^{-4} bis 10^{-6} M liegen. Es ist ebenfalls möglich den Analyten durch eine vorangegangene Flüssigchromatografie zu erhalten⁵.

[5, 14] Anschließend wird die Lösung durch einen entsprechenden Proben-Einlass in eine dünne Kapillare geleitet und in ein starkes elektrisches Feld von ca. $10^6 \text{ V} \cdot \text{m}^{-1}$ gesprüht. Es wird durch ein elektrisches Potential von 3 – 6 kV zwischen der Spitze der Kapillare und einer zylindrischen Gegenelektrode in 0,3 bis 2 cm Entfernung erzeugt.

[9] Die austretende Flüssigkeit wird elektrostatisch aufgeladen und je nach Polung werden positive oder negative Ionen gebildet⁶. Durch die gegensätzlich wirkenden Kräfte der Oberflächenspannung und elektrostatischen Kräfte entsteht ein Taylor-Konus an der Kapillaren-Spitze. Im so genannten *Cone-Jet-Mode* übersteigen die elektrostatischen Kräfte die Oberflächenspannung es bildet sich ein dünner Flüssigkeitsstrahl an der Spitze des Konus in Richtung Gegenelektrode, welcher nach kurzer Zeit in viele kleine geladene Tropfen zerfällt. [5, 13] Diese Initial-Tropfen besitzen einen Überschuss an Ladung. Durch Verdampfen des Lösungsmittels reduziert sich das Volumen der Tropfen wobei die Anzahl der Ladung pro Tropfen gleich bleibt. Schrumpft der Tropfen immer weiter, nimmt das Verhältnis von Ladung zu Oberfläche immer mehr zu. Wenn das *Rayleigh-Limit*⁷ erreicht ist, teilt sich dieser in kleinere Tropfen auf. Da die Tropfen nie eine ideale Kugel formen, sondern leicht

⁵ Bezeichnet als LC-MS oder LC-MS-Kopplung

⁶ Eine positive Spannung erzeugt Kationen und eine negative Anionen.

⁷ Das Rayleigh-limit ist definiert als die maximale Anzahl an Ladungen auf der Tropfenoberfläche, bei der die elektrostatischen Kräfte größer als die Oberflächenspannung auf der Oberfläche eines Tropfens werden. [10]

verformt sind, spalten sich von einem schmalen Ende des Tropfens kleinere ab. Diese *Droplet-Jet-Fission* konnte von der Gruppe um Gomez dokumentiert werden. [15] Oft wird auch der Terminus *Coloumb-Explosion* verwendet. Dieser Prozess zwischen Tropfenabspaltung und Verdampfung des Lösungsmittels wiederholt sich immer wieder, bis Tropfen mit einem Durchmesser von wenigen Nanometern entstehen und nur noch ein Analyt-Molekül enthalten. [5, 16, 13] Wie frei in Gasphase befindliche Ionen aus diesen kleinen Tropfen freigesetzt werden ist noch nicht vollständig geklärt. Es gibt jedoch zwei Modelle: Das *charged residue model* (CRM) postuliert eine schrittweise Verdampfung aller Lösungsmittelmoleküle aus einem Tröpfchen, bis nur noch das Analyt-Molekül übrig bleibt. Für eine Ionisierung sorgen die noch vorhandenen Ionen, wie zum Beispiel Protonen, indem sie auf das Analyt-Molekül übertragen werden. [12, 5] Bei dem *ion evaporation model* (IEM) verdampft das Ion direkt von der Oberfläche des Tröpfchens und geht in die Gasphase über. [17, 5]

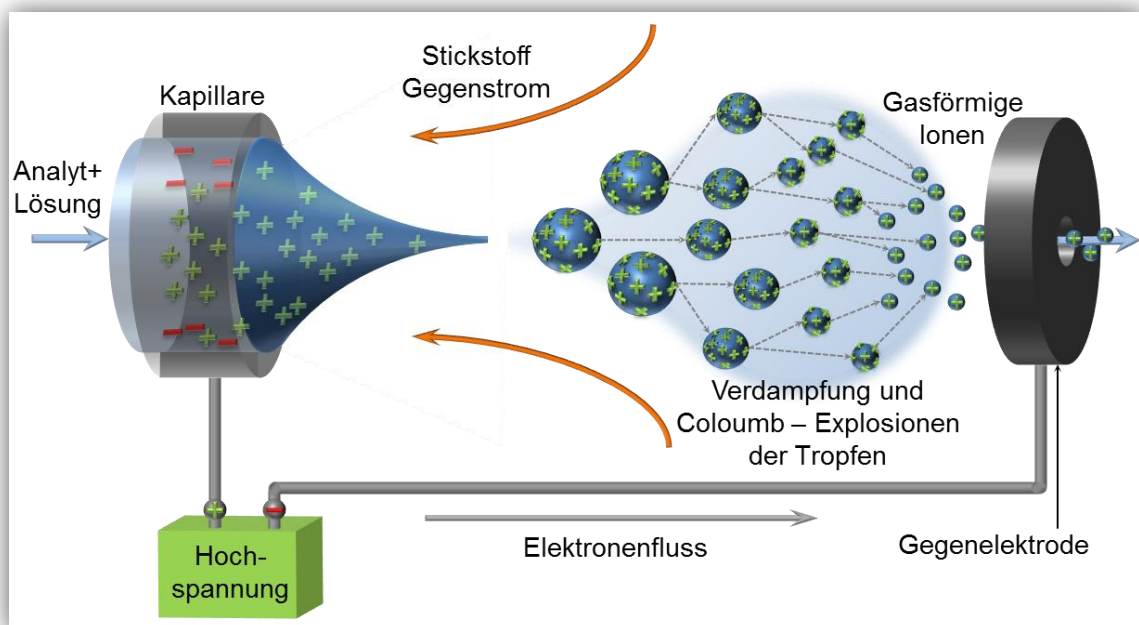


Abbildung 2-3: Schematische Darstellung der Elektrospray-Ionisation

Bei der ESI wird eine Flüssigkeit elektrostatisch aufgeladen sodass sich unter Einfluss eines elektrischen Feldes ein Spray aus kleinen geladenen Tropfen bildet. Nach Reduktion der Tropfengröße durch Verdampfung und Coloumb-Explosionen werden einzelne gasförmige Ionen isoliert und an den Massenanalysator weitergeleitet. Das hier dargestellte Schema entspricht dem Anlegen einer positiven Spannung. (Abbildung nach [16, 5, 13, 9])

Je nach Lösungsmittelzusammensetzung oder Verunreinigungen können nicht nur Protonen für eine Ionisierung des Analyt-Moleküls sorgen, sondern auch andere Ionen

wie Na^+ , K^+ oder NH_4^+ und so verschiedene Addukte bilden. Des Weiteren können auch mehrfach geladene Ionen oder Cluster-Ionen entstehen. Letzteres beschreibt die gemeinsame Ionisierung zweier Analyt-Moleküle. Eine Übersicht der Ionisierungsmöglichkeiten ist in Tabelle 2-1 aufgestellt [5, 16]

Tabelle 2-1: Verschiedene durch ESI erzeugte Ionen

Beispiele für Bildung von positiven und negativen Ionen durch ESI. *M* steht für das Analyt-Molekül, *Alkali* für ein Alkalimetall, *LM* für ein Ion aus einem Lösungsmittelmolekül. (Angepasst nach [5])

Art	Positive Ionen	Negative Ionen
Proton	$[\text{M} + \text{H}]^+$	$[\text{M} - \text{H}]^-$
Alkalimetall	$[\text{M} + \text{Alkali}]^+$ z. B. $[\text{M} + \text{Na}]^+$	
Cluster	$[2\text{M} + \text{H}]^+$ $[2\text{M} + \text{Alkali}]^+$	$[2\text{M} - \text{H}]^-$
Addukte	$[\text{M} + \text{LM} + \text{H}]^+$ $[\text{M} + \text{LM} + \text{Alkali}]^+$ z. B. $[\text{M} + \text{NH}_4]^+$	$[\text{M} + \text{LM} + \text{H}]^-$

Da die ESI unter Atmosphärendruck stattfindet, der Massenanalysator und Detektor aber unter Vakuum arbeiten, müssen die produzierten Ionen anschließend durch verschiedene Pumpstufen transportiert werden bis der benötigte Unterdruck erzeugt wurde. [9]

2.2.3 Massenanalysatoren

Die hauptsächliche Aufgabe von Massenanalysatoren besteht darin, die ionisierten Moleküle nach ihrem spezifischen m/z – Werten aufzutrennen. Damit die gasförmigen Ionen auf ihrer Flugbahn nicht zufällig mit anderen Neutralteilchen oder Ionen kollidieren und in Folge dessen fragmentieren, arbeiten alle Massenanalysatoren unter Hochvakuum bei $10^{-4} - 10^{-7} \text{ Pa}$ (vgl. Abbildung 2-1). [8]

Es gibt verschiedene physikalische Prinzipien, mit denen man eine Massentrennung erreichen kann. In Tabelle 2-2 sind einige wichtige Eigenschaften der im Folgenden vorgestellten Massenanalysatoren zusammengefasst. Da die für diese Arbeit verwendeten Daten mit einem FT-ICR Massenspektrometer erzeugt wurden, wird auf diese Methode etwas näher eingegangen.

Tabelle 2-2: Zusammenfassung der Eigenschaften von Massenanalysatoren

In dieser Tabelle sind die wichtigsten Kennzahlen und Eigenschaften einiger Massenanalysatoren dargestellt. (angepasst nach [9])

	Q	QIT	TOF	B	FT-ICR	Orbitrap
obere Massengrenze¹	4.000	6.000	>1.000.000	20.000	30.000	50.000
Auflösung²	2.000	4.000	5.000	100.000	>500.000	>100.000
Genauigkeit³	100	100	200	<10	<5	<5
Probenzufuhr	kont.	gepulst	gepulst	kont.	gepulst	gepulst
Druck⁴	10^{-5}	10^{-3}	10^{-5}	10^{-6}	10^{-10}	10^{10}
Tandem MS	ja	ja		ja	ja	
1: in Dalton, 2: FWHM bei m/z 1000, 3: in ppm, 4: in Torr, kont. = kontinuierlich						

2.2.3.1 Flugzeit Massenanalysator

Werden Ionen mit unterschiedlichen m/z - Werten zur gleichen Zeit aus einer gepulsten Ionenquelle in eine feldfreie Driftstrecke geschossen, trennen sich die Ionen entlang der Flugstrecke zeitlich auf. Die führt dazu, dass Ionen mit geringerem m/z eher am Ende der Flugröhre, und damit am Detektor ankommen als Ionen mit höherem m/z . Ermöglicht wird dies durch die unterschiedlichen Geschwindigkeiten auf die Ionen unterschiedlicher Masse in einem elektrischem Feld beschleunigt werden. Von diesen *Time-Of-Flight* oder *Flugzeit-Analysatoren* (TOF) wurden verschiedene Varianten entwickelt. Die einfachste und erste ist der lineare TOF-Analysator. Ionen werden hier mit Hilfe von MALDI erzeugt und bewegen sich von der Ionisationsquelle bis hin zum Detektor auf einer geraden Bahn. [5, 9]

Beim *Reflektron* bzw. dem *Reflektor-TOF-Analysator* wird ein statisches elektrisches Feld genutzt um die Flugrichtung der Ionen umzukehren. Dieser Ionenspiegel ist linear zur Ionenquelle angebracht. Die Ionen verlieren beim Eintritt in das elektrische Feld ihre kinetische Energie und werden erneut beschleunigt, wenn sie es verlassen. Werden die Ionen in einem Winkel bezüglich des Ionenstrahls von der Ionenquelle ausgeworfen, kann der Detektor direkt neben der Ionenquelle platziert werden. Der Einsatz eines Reflektrons hat den Vorteil, dass die Ionen in ihrer Flugzeit korrigiert werden und somit das Auflösungsvermögen erhöht wird. [5, 10]

Um auch andere Ionenquellen außer MALDI mit TOF-Analysatoren koppeln zu können wurde die *TOF mit orthogonaler Beschleunigung* (oaTOF) entwickelt. Mit dieser Methode werden aus einem kontinuierlichem Ionenstrahl Ionenpakete durch Pulse

orthogonal abgelenkt. Anschließend treten diese wieder in die Flugröhre, mit oder ohne Reflektron, ein. Mit dieser Methode können sehr viele Spektren in sehr kurzer Zeit gemessen werden. [5]

2.2.3.2 Magnetsektor-Analysatoren

Bewegen sich geladene Teilchen senkrecht zu einem Magnetfeld wirkt die so genannte Lorentz-Kraft im rechten Winkel zur Bewegungsrichtung und Magnetfeld auf die Teilchen. Dadurch werden Ionen eines bestimmten m/z - Werts auf eine Kreisbahn mit charakteristischem Radius abgelenkt. Wird nun die Stärke des Magnetfeldes verändert, können Ionen unterschiedlicher m/z - Werte an einem Punkt detektiert werden. Die Voraussetzung für eine gute Auflösung ist, dass alle Ionen gleichen m/z - Wertes auch die gleiche kinetische Energie besitzen. Da diese aber oft streuen und das magnetische Feld eine Energiedispersion hervorruft, werden zusätzlich zu den Magnetsektoren⁸ elektrostatische Sektoren⁹ eingesetzt. Diese Kombination führt zu einer doppelten Fokussierung der Ionen, nach Energie und Richtung, und damit zu einer Verbesserung der Auflösung und Massenbestimmung. [8]

2.2.3.3 Quadrupol-Analysatoren und Quadrupol-Ionenfallen

Ein *Quadrupol* (Q) besteht aus vier quadratisch angeordneten zylindrischen oder hyperbolischen Elektroden in deren Zentrum ein elektrisches Feld erzeugt wird. Dafür werden an jeweils gegenüberliegenden Elektroden eine Spannung angelegt, die sich aus zwei Komponenten zusammensetzt: einer Gleichspannung (DC) und einer Wechselspannung (AC) im Radiofrequenz (RF) – Bereich. Treten Ionen in den Quadrupol ein, werden sie von der Elektrode mit entgegengesetzter Ladung angezogen. Bevor es zu einer Entladung kommt, wechselt die Polung der Elektroden und das Ion wird nun abgestoßen und von einer anderen Elektrode angezogen. Ionen eines bestimmten m/z - Verhältnisses werden so auf stabilen Trajektorien durch den Quadrupol geleitet und bei Austritt aus dem Quadrupol detektiert. Durch Veränderung der DC – und RF – Spannung mit gleichbleibendem DC/RF - Verhältnis wird über einen bestimmten m/z - Bereich gescannt. Ein nachgeschalteter Detektor misst die Ionen und ein Massenspektrum wird erzeugt. Durch ihren einfachen Aufbau sind Quadrupol-Analysatoren im Vergleich zu anderen Analysatoren günstig, leicht und klein. Sie weisen eine hohe Transmission und hohe Scangeschwindigkeit auf. [5, 16]

⁸ Symbol: B

⁹ Symbol: E

Quadrupole können nicht nur als Massenanalysator eingesetzt werden, sondern sind auch sehr gut als Ionentunnel zwischen zwei Komponenten des Massenspektrometers oder als Kollisionszelle für Stoßexperimente geeignet. Bei diesen Anwendungen wird die RF-Spannung verwendet und die Gleichspannung ist gleich null. Des Weiteren existieren noch Hexapole und Oktapole, die vom Aufbau her ähnlich wie Quadrupole sind, jedoch aus sechs bzw. acht Elektroden bestehen. Sie werden ebenfalls als Ionentunnel verwendet. Ihre Massenspezifität ist zwar geringer, dafür treten aber geringere Verluste bei den zu überführenden Ionen auf als bei Quadrupolen. [9]

Von einer *Quadrupol-Ionenfalle* spricht man, wenn ein oszillierendes elektrisches Feld genutzt wird um Ionen zu speichern. Davon gibt es zwei Typen: Die *lineare Quadrupol-Ionenfalle* (LIT), bei der an den Enden des Quadrupols zwei weitere Elektroden angebracht werden um die Ionen zu reflektieren, und die *dreidimensionale Quadrupol-Ionenfalle* (QIT), welche ellipsoide Randelektroden und anstatt vier einzelner Elektroden eine Ringelektrode nutzt. Ionen verschiedener m/z - Werte werden so in einem zwei- oder dreidimensionalen Quadrupol-Feld gehalten und durch Veränderung der RF-Spannung detektiert. [8]

2.2.3.4 Orbitrap

Erst seit dem Jahr 2005 sind Massenspektrometer mit Orbitrap-Analysator erhältlich. Das Funktionsprinzip ähnelt in einigen Punkten der FT-ICR: Es werden oszillierende Ionenbewegungen mit Hilfe der Bildstromdetektion aufgenommen und anschließend durch eine Fourier-Transformation in entsprechende Frequenz – und Intensitätssignale umgewandelt. Der Orbitrap nutzt jedoch anstatt eines magnetischen ein elektrostatisches Feld um Ionen einzufangen und auf periodische Trajektorien zu lenken. Er ist aus einer spindelförmigen zentralen Elektrode und einer zweiteiligen äußeren fassförmigen Elektrode aufgebaut. An die innere Elektrode wird eine Spannung angelegt, sodass die Ionen von ihr angezogen werden. Die äußere Elektrode wird auf Grundpotential gehalten. Es entsteht ein quadrupol – logarithmisches Feld in dem die Ionen, durch ein Gleichgewicht zwischen der elektrostatischen Kraft und der Zentrifugalkraft, auf spiralförmige oszillierende Trajektorien gelenkt werden. Da die Frequenz der Ionen abhängig von ihrem m/z - Wert ist, ergibt die Messung des durch die Ionenbewegung induzierten Stroms und die Umwandlung der Signale durch Fourier-Transformation in die Frequenz-Domäne ein Massenspektrum. Um bessere Ergebnisse zu erhalten, wird meist eine so genannte C-Trap genutzt um die Ionen vor der Injektion in den Orbitrap zu sammeln und abzukühlen. Vorteile gegenüber der FT-ICR sind die fehlenden starken Magneten und

der dadurch geringere Platzbedarf und geringerer Preis. Des Weiteren erreichen Orbitrap-Analysatoren bei höheren m/z - Werten bessere Auflösungen. [5, 8]

2.2.3.5 Fourier–Transform Ion–Cyclotron–Resonance

Jeder Massenanalysator hat seine Vor- und Nachteile bei der Analyse bestimmter Stoffgruppen oder Gemischen. Sind die Moleküle sehr groß und ist die Genauigkeit sowie Auflösung von großer Bedeutung, kommt oft die *Fourier-Transform Ion-Cyclotron-Resonance* Massenspektrometrie, kurz *FT-ICR MS*, zum Einsatz. Erstmals wurde diese Technik von Comisarov und Marshall im Jahr 1974 vorgestellt. [18]

Es wird eine Kombination aus magnetischen und elektrischen Feldern genutzt um Ionen einzufangen, auf eine Kreisbahn zu zwingen, anzuregen und anschließend mit Hilfe einer Bildstrommessung die Ionenströme, d.h. die Frequenzen der Ionen, zu detektieren. Eine Fourier-Transformation konvertiert anschließend die gemessenen Daten von der Zeit- in die Frequenzdomäne. Die einzelnen Schritte werden hier nun näher erläutert.

Für eine ICR-Zelle benötigt man ein starkes homogenes Magnetfeld, welches durch einen supraleitenden Magneten in z-Richtung erzeugt wird, zwei gegenüberliegende Trapping-Elektroden an den Enden die ein elektrisches Feld erzeugen damit die Ionen in der ICR-Zelle gehalten werden und jeweils zwei gegenüberliegende Elektroden für die Anregung der Ionen durch ein elektrisches Feld und die Detektion der Bildströme der Ionen. In Abbildung 2-4 sind zwei mögliche Architekturen abgebildet mit den jeweiligen Elektroden und den Koordinaten zur Orientierung. Die Ionen treten in z-Richtung in die Zelle ein.

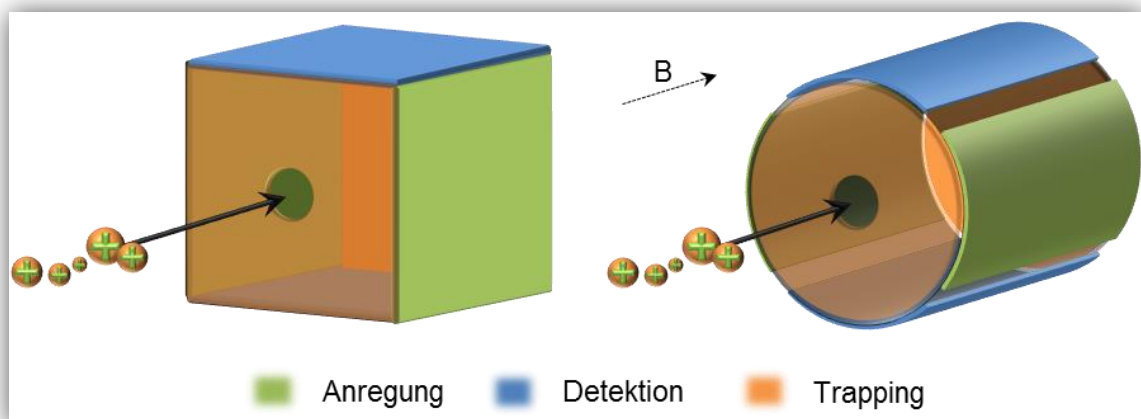


Abbildung 2-4: Beispiele für die Architektur einer ICR-Zelle

Links: Eine kubische ICR-Zelle. **Rechts:** Eine zylindrische ICR-Zelle. Werden segmentierte Trapping-Platten eingesetzt, sodass diese zusätzlich zum Einfangen der Ionen auch für die Anregung genutzt werden können, handelt es sich um eine *Infinity Cell*[™]. (angepasst nach [5, 19])

Ein geladenes Teilchen mit der Masse m und der Ladung q , welches sich mit einer konstanten Geschwindigkeit v senkrecht zu einem homogenen Magnetfeld B bewegt wird durch die Lorentz-Kraft F_L auf eine Kreisbahn mit dem Radius r gezwungen.

$$F_L = qvB \quad (2.1)$$

Diese Ionenbewegung ist in Abbildung 2-5 für positive und negative Ionen schematisch dargestellt. Die Lorentz-Kraft, die hier als Zentripetalkraft wirkt, steht im Gleichgewicht mit der Zentrifugalkraft F_Z die sich aus der Masse m und der Winkelbeschleunigung des Ions ergibt:

$$F_Z = m \cdot \frac{\delta v}{\delta t} = m \cdot \frac{v^2}{r} \quad (2.2)$$

$$F_Z = F_L \rightarrow m \cdot \frac{v^2}{r} = qvB \quad (2.3)$$

Daraus ergibt sich ein Ionen Cyclotron Radius der direkt proportional zum m/z - Verhältnis des Ions und dessen Geschwindigkeit ist. Je stärker das Magnetfeld, desto kleiner werden die Radien. So können auch hochmolekulare Moleküle mit m/z 10.000 und mehr in relativ kleinen ICR-Zellen analysiert werden. Das rechte Diagramm in Abbildung 2-6 zeigt den Radius in Abhängigkeit des m/z - Wertes bei verschiedenen Magnetfeldstärken.

$$r = \frac{mv}{qB} \quad (2.4)$$

Die Winkelgeschwindigkeit ist definiert als $\omega = v/r$. Durch Einsetzen in Formel 2.3 und Umstellung nach ω ergibt sich die Cyclotron-Winkelfrequenz:

$$\omega_c = \frac{qB}{m} \quad (2.5)$$

Die ungestörte Cyclotron-Frequenz $f_c = \omega_c/2\pi$ ist damit definiert als

$$f_c = \frac{qB}{2\pi \cdot m} \quad (2.6)$$

Das bedeutet, dass alle Ionen mit einem bestimmten m/z - Wert die gleiche Cyclotron-Frequenz unabhängig von deren Geschwindigkeit besitzen. Es ist also nicht wichtig die Ionen in ihrer Geschwindigkeit zu fokussieren wie bei anderen Massenanalysatoren.

[5, 9, 18]

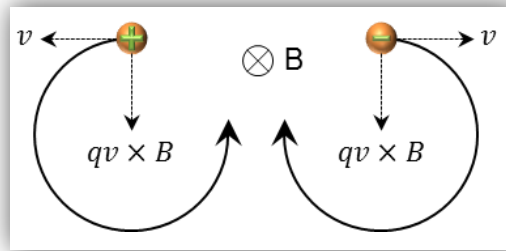


Abbildung 2-5: Kreisbewegung eines Ions in einem Magnetfeld

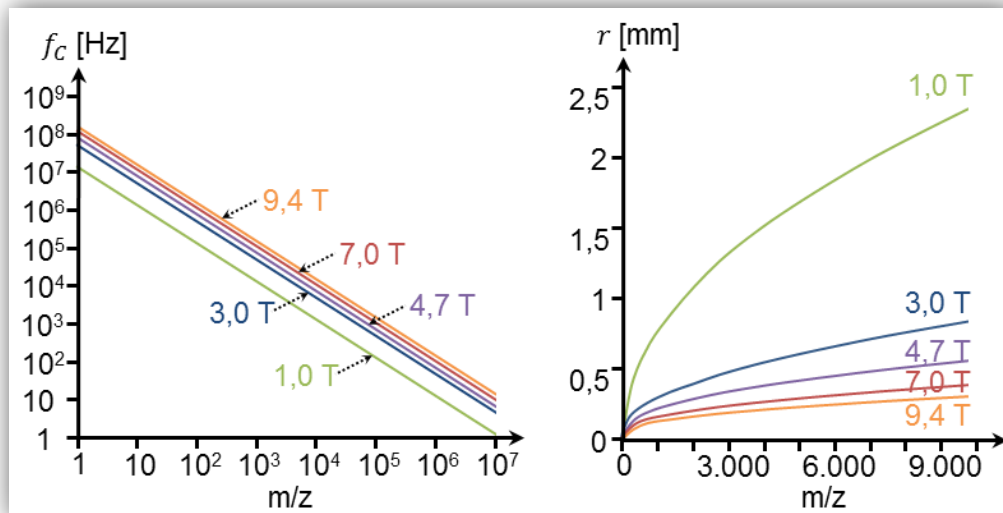
Auf ein Ion mit Geschwindigkeit v und Ladung q , dessen Bewegungsrichtung orthogonal zum Magnetfeld B verläuft, wirkt eine Kraft durch die es auf eine Kreisbahn mit bestimmtem Radius gezwungen wird.

(angepasst nach [19])

Je nach m/z - Wert der zu messenden Ionen muss ein FT-ICR-Massenspektrometer Frequenzen im Bereich von kHz bis MHz abdecken. Zum Beispiel besitzt ein einfach geladenes Ion mit m/z 50 in einem Magnetfeld der Stärke $B = 9,4 \text{ T}$ eine Cyclotron-Frequenz von $f_c = 3 \text{ MHz}$ wie in Formel 2.7 dargestellt¹⁰. Bei m/z 5000 beträgt die Frequenz nur noch $f_c = 300 \text{ kHz}$. In Abbildung 2-6 ist der Zusammenhang zwischen dem m/z - Verhältnis und der Cyclotron-Frequenz für verschiedene Magnetfeldstärken dargestellt.

$$f_c = \frac{(1 \cdot 1.602176462 \cdot 10^{-19} \text{ C}) \cdot 9,4 \text{ T}}{2\pi \cdot (50 \cdot 1.66054 \cdot 10^{-27} \text{ kg})} \approx 3 \text{ MHz} \quad (2.7)$$

Heutige Geräte benutzen Magneten mit Feldstärken von 7 T bis 15 T . [5, 19]

Abbildung 2-6: Cyclotron-Frequenz und -Radius in Abhängigkeit vom m/z - Wert

Das linke Diagramm zeigt die Cyclotron-Frequenz f_c und das rechte Diagramm den Cyclotron-Radius r in Abhängigkeit des m/z - Wertes der Ionen bei verschiedenen Magnetfeldstärken von 1 bis $9,4 \text{ T}$.

(angepasst nach [19])

¹⁰ Der Wert von $1.602176462 \cdot 10^{-19} \text{ C}$ entspricht der Elementarladung e [41] und $1.66054 \cdot 10^{-27} \text{ kg}$ der Masse einer Atomaren Masseneinheit u [27].

Die Cyclotron-Bewegung hält die Ionen auf einer Kreisbahn mit konstantem Radius unter der Voraussetzung, dass keine Kollisionen mit Neutralteilchen oder anderen Ionen auftreten und die thermische Energie stabil bleibt. Sie ist jedoch nicht zur Detektion der Ionen nutzbar. Deshalb werden die Ionen mit Hilfe eines elektrischen Potentials angeregt, sodass sie ihren Radius vergrößern. Das elektrische Feld sollte dabei senkrecht zum Magnetfeld ausgerichtet sein. Es oszilliert sinusförmig mit der Zeit t nach Formel 2.8 bei derselben Frequenz wie die Cyclotron-Frequenz ω_c der Ionen.

$$E(t) = E_0 \cdot \cos \omega_c t ; E_0 = \frac{2V_0}{d} \quad (2.8)$$

Wobei E_0 das elektrische Feld zwischen zwei parallelen Elektroden mit der Entfernung d zueinander. An den Elektroden liegen jeweils $+V_0$ und $-V_0$ Volt an. Durch gleiche Richtung und Frequenz des elektrischen Feldes wie die der Ionen absorbieren sie Energie und werden beschleunigt, d.h. der Radius ihrer Flugbahn vergrößert sich und sie bewegen sich spiralförmig nach außen (vgl. Abbildung 2-7 links).

$$A(t) = \text{Kraft} \cdot \text{Geschwindigkeit} = qE(t) \cdot v \quad (2.9)$$

Der Radius r von Ionen nach einer bestimmten Anregungsdauer T_{Anregung} kann aus Formel 2.10 hergeleitet werden und errechnet sich aus

$$r = \frac{E_0 \cdot T_{\text{Anregung}}}{2B} = \frac{V_0 \cdot T_{\text{Anregung}}}{d \cdot B} \quad (2.10)$$

Nach der Anregung der Ionen ist der Radius nun nicht mehr von den m/z - Werten der Ionen wie vor der Anregung, vgl. Formel 2.4, abhängig. Dies bringt den großen Vorteil mit sich, dass alle Ionen eines bestimmten m/z - Bereichs auf einen Radius beschleunigt werden können. [19]

Die Anregung von Ionen kann für verschiedene Zwecke genutzt werden: Zum einen können damit Ionen aus der ICR-Zelle entfernt werden, wenn der Radius nach der Anregung größer als der Radius der ICR-Zelle ist. Zum anderen wird die Anregung genutzt um Ionen im Anschluss detektieren zu können. Des Weiteren kann die kinetische Energie von Ionen so stark erhöht werden, sodass sie dissoziieren und mehrere Tochter-Ionen bilden. Dieses Prinzip nennt man *collision induced dissociation* (CID) und wird später noch erläutert. [19, 5]

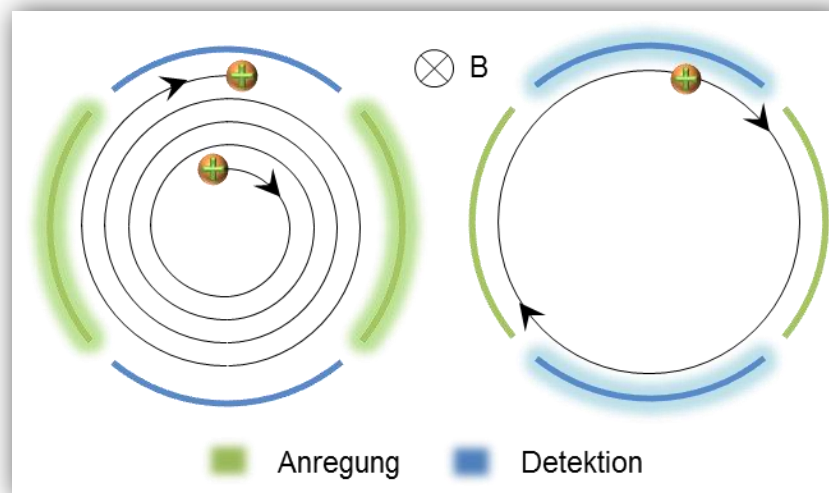


Abbildung 2-7: Schema der Trajektorie von Ionen bei Anregung und Detektion

Links: Liegt ein RF-Potenzial mit gleicher Frequenz wie die Cyclotron-Frequenz an den Anregungs-Elektroden an, bewegen sich die Ionen spiralförmig nach außen auf größere Radien. **Rechts:** Nach der Anregung erfolgt die Detektion der Bildströme. Anregung und Detektion erfolgen zeitlich getrennt.

(angepasst nach [19])

Je nachdem, ob nur Ionen mit einem bestimmten m/z - Wert gemessen werden sollen oder alle Ionen in der ICR-Zelle, existieren verschiedene Anregungsmodi. Die Anregung einzelner m/z - Werte erreicht man durch Rechteckpulse, wo für eine bestimmte Zeit eine bestimmte Frequenz genutzt wird. Sollen mehrere oder sogar alle Ionen gemessen werden, kann die *chirp*-Anregung oder *stored waveform inverse fourier transformation* (SWIFT) genutzt werden. Ein Chirp regt die Ionen verschiedener m/z - Werte nacheinander an, indem alle benötigten Frequenzen nacheinander abgetastet werden. Bei der SWIFT wird aus den m/z - Werten, die gemessen werden sollen, ein Anregungsprofil in der Frequenz-Domäne erstellt. Dieses wird durch inverse FT in die Zeit-Domäne übersetzt und durch einen RF-Generator erzeugt. Diese Methode ermöglicht die genaue Anpassung der Anregungswellenform an das jeweilige Experiment und ist dadurch sehr genau. Mit dieser Methode können auch Ionen aus der ICR-Zelle entfernt werden. Dadurch ist es möglich nur Ionen eines bestimmten m/z - Wertes in der ICR-Zelle zu belassen und zum Beispiel Fragmentierungs-experimente durchzuführen. [5, 20, 21]

Die bisher gemachten Annahmen über Ionenbewegung und Anregung setzen voraus, dass sich die Ionen in einem unendlich ausgedehnten magnetischen- bzw. elektrischen Feld bewegen. Da das Magnetfeld nur in der xy-Ebene begrenzend wirkt und das elektrische Feld nur in y-Ebene Einwirkung auf die Ionen hat, können sie sich in z-Richtung frei bewegen. Daher benötigt man, ähnlich wie bei den Quadrupol-Zellen ein

Trapping-Potenzial welches durch Elektroden an den Enden der ICR-Zelle erzeugt wird. Die Ionen werden reflektiert und im Inneren der Zelle gehalten. Dafür wird zum Beispiel ein drei-dimensionales quadrupolares Feld benutzt. Dieses übt auf die Ionen eine radiale Kraft

$$F_R = qE(r) = \frac{qV_{trap}\alpha}{a^2}r \quad (2.11)$$

aus, die entgegengesetzt zur Lorentz-Kraft wirkt. Die Gesamtkraft F_{Gesamt} der die Ionen ausgesetzt sind ergibt sich somit aus einer Kombination von Formel 2.3 mit Formel 2.11 zu

$$F_{Gesamt} = F_L - F_R = m\omega^2r = qB\omega r - \frac{qV_{trap}\alpha}{a^2}r. \quad (2.12)$$

Nach Umformung ergibt sich die quadratische Gleichung

$$\omega^2 - \frac{qB}{m}\omega + \frac{qV_{trap}\alpha}{ma^2} = 0 \quad (2.13)$$

Mit den Lösungen

$$\omega_+ = \frac{\omega_C}{2} + \sqrt{\left(\frac{\omega_C}{2}\right)^2 - \frac{\omega_Z}{2}} \text{ und} \quad (2.14)$$

$$\omega_- = \frac{\omega_C}{2} - \sqrt{\left(\frac{\omega_C}{2}\right)^2 - \frac{\omega_Z}{2}} \quad (2.15)$$

Wobei ω_C die ungestörte Cyclotron-Frequenz nach Formel 2.5 bzw. Formel 2.6 und

$$\omega_Z = \sqrt{\frac{2qV_{trap}\alpha}{ma^2}} \quad (2.16)$$

die Schwingungsfrequenz entlang der z-Achse, oder oszillierende Trapping-Frequenz, beschreibt. Die Gesamtbewegung eines Ions in einer ICR-Zelle ergibt sich also aus dem Zusammenspiel der reduzierten Cyclotron-Frequenz ω_+ , der Magnetron-Frequenz ω_- und der oszillierenden Trapping-Frequenz ω_Z . Diese drei Bewegungen und die Gesamtbewegung ist in Abbildung 2-8 dargestellt. **[5, 19]** Die Magnetron- und Trapping-Frequenz ist sehr viel kleiner als die reduzierte Cyclotron-Frequenz und werden daher nicht gemessen. Sie können jedoch als schmale Seitenbanden im Massenspektrum auftreten, wenn zum Beispiel die Magnetachse nicht mit der Achse der Ionenfalle genau übereinstimmt. **[19]**

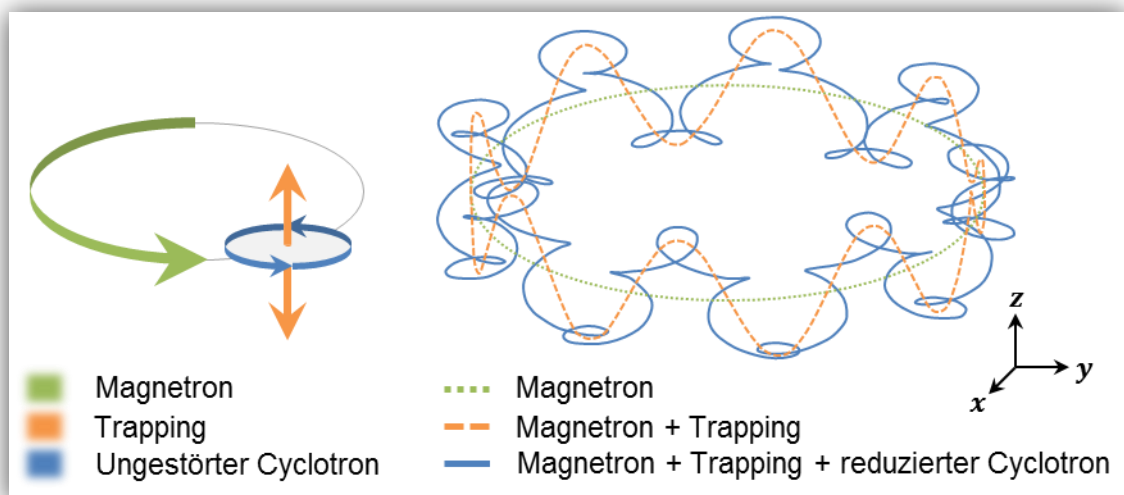


Abbildung 2-8: Ionen-Bewegungen in einer ICR-Zelle

Es sind die drei natürlichen Bewegungen der Ionen schematisch dargestellt. Das Magnetfeld zeigt in z-Richtung. **Links:** Separierte Trajektorien von Magnetron-, Trapping- und ungestörter Cyclotron-Bewegung. **Rechts:** Gesamtbewegung eines Ions in einer ICR-Zelle mit Magnetfeld B und quadropolarem Trapping-Potenzial. Sie setzt sich zusammen aus Magnetron-, Trapping- und reduzierter Cyclotron-Bewegung. (angepasst nach [5])

Für die Detektion werden die Ionen auf einen Radius der etwas kleiner ist als der Radius der ICR-Zelle angeregt bzw. beschleunigt. Da sich Ionen gleichen m/z - Wertes durch die resonante Anregung als kohärente Ionenpakete¹¹ bewegen, kann zwischen den Detektorplatten ein induzierter Strom gemessen werden. Dieser Bildstrom wird verstärkt, digitalisiert und aufgezeichnet. Das komplette Signal¹² besteht aus der Überlagerung aller Frequenzen aller Ionenpakete mit unterschiedlichen m/z - Werten. Mit der Zeit wird das Signal durch Dämpfung und durch Kollisionen mit Restgas schwächer. Um die Anzahl der Kollisionen so gering wie möglich zu halten damit das Signal solange wie möglich erhalten bleibt, wird die ICR-Zelle unter Ultrahochvakuum bei 10^{-8} bis 10^{-10} mbar gehalten werden. Die Amplituden entsprechen dabei den Häufigkeiten der Ionen. Um die einzelnen Frequenzen zu trennen, wird der Transient mit Hilfe der Fourier Transformation von der Zeit - Domäne in die Frequenz - Domäne dekonvuliert. Es entstehen Peaks mit der Form einer Lorentz-Kurve, deren Breite durch die Auflösung und Höhe durch die Amplitude bestimmt wird. Durch Umstellung von Formel 2.5 kann der m/z - Wert eines Peaks ermittelt werden. In Abbildung 2-9 ist dieser Vorgang schematisch dargestellt. [19, 5]

¹¹ Die Ionen sind an einem Punkt auf ihrer Kreisbahn konzentriert und nicht verteilt.

¹² Es werden auch die Bezeichnungen Free Induction Decay (FID) oder Transient verwendet.

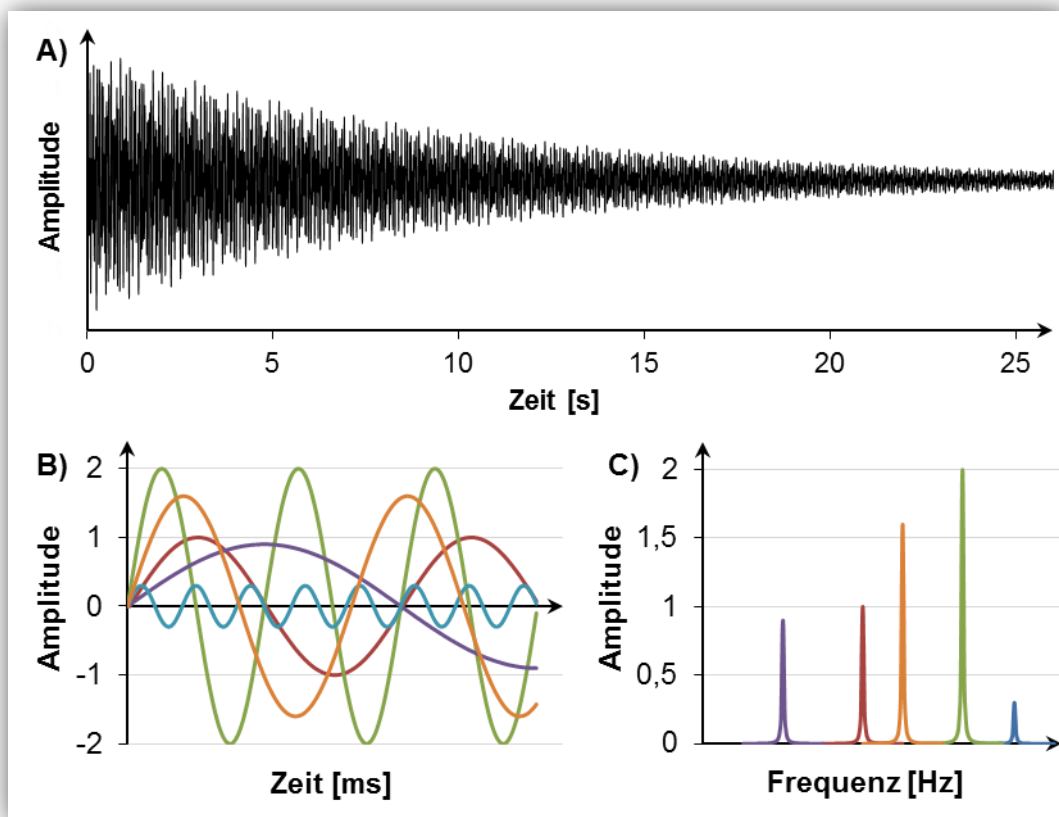


Abbildung 2-9: Umwandlung des Transienten von der Zeit- in die Frequenzdomäne

A) Ein Transient besteht aus einer Überlagerung aller **B)** Einzelschwingungen von Ionenpaketen unterschiedlicher m/z - Werte. Die Amplitude der Einzelschwingungen entspricht der absoluten Häufigkeit von Ionen eines m/z - Wertes. **C)** Durch Anwendung der Fourier-Transformation wird der Transient von der Zeitdomäne in die Frequenzdomäne umgewandelt. Durch Umrechnung der einzelnen Frequenzen in die jeweiligen m/z - Werte erhält man ein typisches Massenspektrum. Die Peaks haben die Form einer Lorentz-Kurve. (angepasst nach [20])

Die Beobachtungsdauer des Transienten beträgt zwischen 0,5 und 30 s. Je kürzer die Messzeiten, desto weniger Zyklen der Ionen werden gemessen und desto geringer die Auflösung. In der Frequenzdomäne zeigt sich dies durch so genannte Seitenbanden, oder Wiggles, wie in Abbildung 2-10 dargestellt. Aus diesem Grund werden vor der Fourier-Transformation verschiedene Optimierungsalgorithmen auf den Transienten angewandt: Bei der *Apodisierung* wird das Signal mit einer periodischen Funktion, wie zum Beispiel $f(t) = \sin(t)$, multipliziert, um eine Dämpfung bis auf den Null zu simulieren. Das *Zerofilling* wird verwendet, um die Peaks in der Frequenz-Domäne zu glätten. Dafür werden so viele oder doppelt so viele Nullen am Ende des Transienten angehängt wie es Datenpunkte gibt. [20]

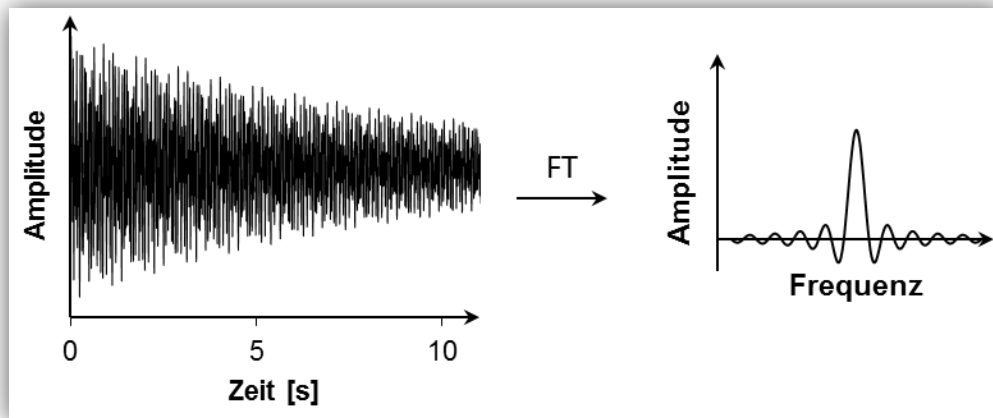


Abbildung 2-10: Auswirkung einer kurzen Messzeit auf die Frequenzdomäne

Durch kürzere Messzeiten wird der Transient nicht komplett aufgenommen. Durch Fourier-Transformation entsteht keine Lorentz-Kurve sondern eine Kombination aus einer *sinc* – Funktion und einer Lorentz-Kurve. (angepasst aus [5])

Ein Beispiel eines FT-ICR-Massenspektrometers ist in Abbildung 2-11 dargestellt. Es handelt sich um ein Schema aus der Bruker Daltonics solarix Serie.

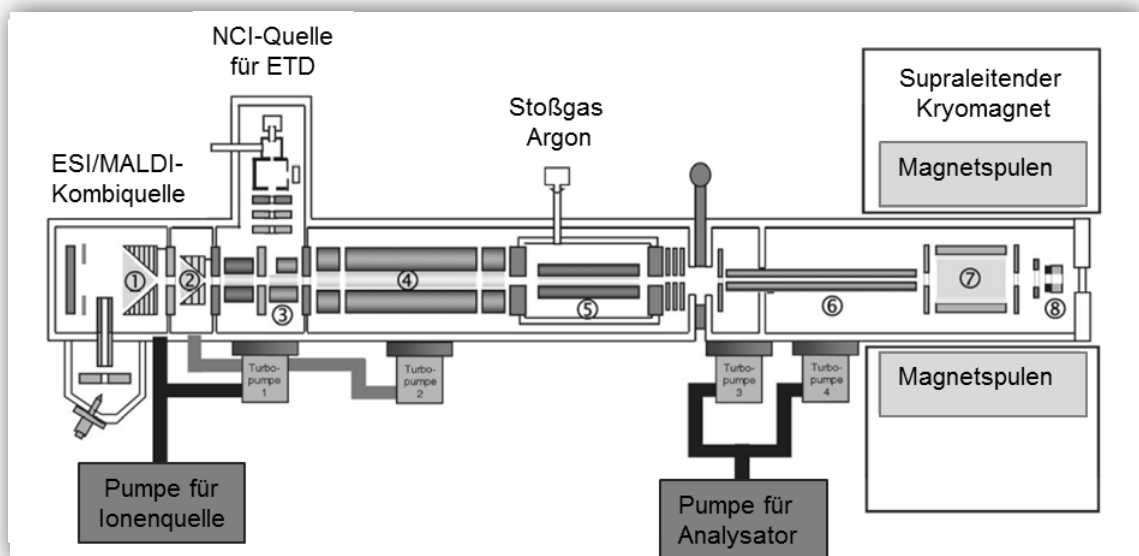


Abbildung 2-11: Schema des Bruker Daltonics Solarix FT-ICR-MS

Die Ionenquelle ist eine ESI/MALDI-Kombiquelle. Die Ionen werden durch einen Ionentrichter (1,2) in einen RF-Hexapol geleitet und fokussiert (3). Anschließend werden sie in einen Quadrupol weitergeleitet (4) der Ionen selektiert oder als Breitband alle Ionen durchlässt. Ein nachgeschalteter Quadrupol (5) kann nun die Ionen sammeln, stoßaktivieren oder für ETD mit Radikal-Ionen reagieren lassen. Danach gelangen sie durch einen Hexapol (6) in die ICR-Zelle (7) und werden dort detektiert. Es existiert eine Hohlkathode für ECD (8). (angepasst aus [5])

2.3 Das Massenspektrum

2.3.1 Allgemeine Beschreibung

Als Massenspektrum bezeichnet man eine zweidimensionale grafische oder tabellarische Darstellung von m/z - Werten in Abhängigkeit von der Signalintensität. Auf der Abszisse werden die m/z - Werte und auf der Ordinate die Intensitäten, die den jeweiligen Häufigkeiten der Ionen im Analyten widerspiegeln, aufgetragen. Es gibt zwei verschiedene grafische Darstellungsformen: das Profilspektrum und das Linienspektrum welche in Abbildung 2-12 Links und Mitte schematisch dargestellt sind. Ein Profilspektrum erhält man zum Beispiel bei der FT-ICR nach der Fourier-Transformation von der Zeit- in die Frequenzdomäne. Die einzelnen Signale werden als Peaks bezeichnet und zeigen an, dass der Analyt Ionen dieses m/z - Wertes enthält. Falls Fragmentierungs-Experimente durchgeführt werden, kann ein Peak auch ein Bruchstück eines Ions repräsentieren. Als *Basis-Peak* wird der höchste Peak in einem Massenspektrum genutzt. Anhand der Intensität dieses Peaks kann das Massenspektrum auf die relative Intensität genormt werden. In einem experimentellen Massenspektrum tritt immer ein Rauschen auf. Je nachdem wie stark das Rauschen ist, können Peaks mit sehr geringer Intensität im Rauschen untergehen. Ein Linienspektrum stellt jeden Peaks als diskretes Wertepaar in einem Histogramm dar. Die jeweiligen Intensitäten ergeben sich aus den Flächen eines Peaks und die m/z -Werte stellen den Schwerpunkt des Peaks dar. [5]

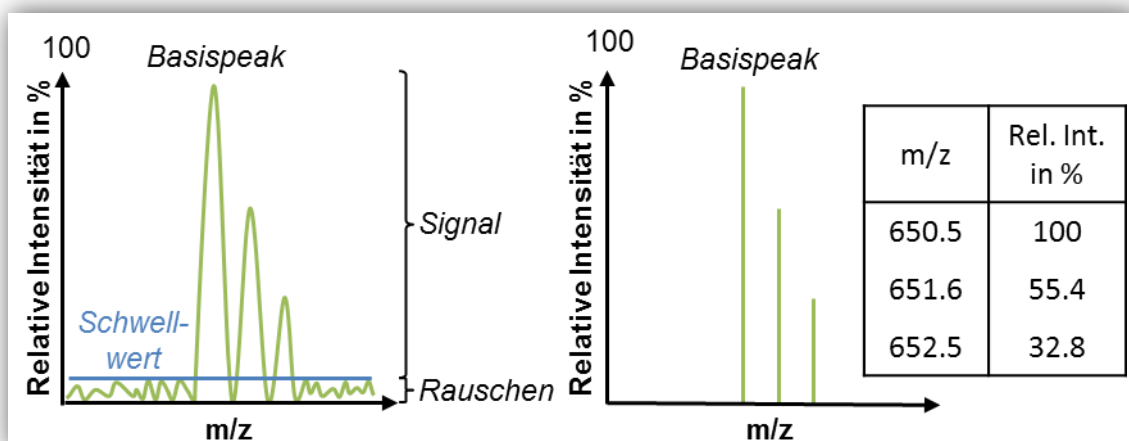


Abbildung 2-12: schematische Darstellung eines Massenspektrums

Bei einem Massenspektrum werden die relativen oder absoluten Intensitäten gegen den m/z -Wert aufgetragen. Der höchste Peak wird als Basispeak bezeichnet. **Links:** Ein Profilspektrum mit drei Peaks.

Alle Messdaten unterhalb eines bestimmten Schwellwertes werden als Rauschen bezeichnet.

Rechts: Tabellarische Darstellung des Linienspektrums. (erstellt nach [5])

Nicht jeder Peak in einem Massenspektrum repräsentiert ein bestimmtes Molekül, welches im Analyt enthalten ist. Wie später noch erläutert wird, gehören zu einer Verbindung mehrere aufeinanderfolgende Peaks aufgrund der Isotopenzusammensetzung der Elemente. Diese besitzen unterschiedliche Intensitäten und sind berechenbar. Der Peak, der dem Molekulargewicht eines Moleküls entspricht, wird auch als *Molekül-Ionen-Peak* bezeichnet.

2.3.2 Signal-zu-Rausch-Verhältnis

Wie in Abbildung 2-12 Links dargestellt, besitzen Massenspektren ein Rauschen, welches zum Beispiel durch die Elektronik im Massenspektrometer verursacht werden kann. Das Signal-zu-Rausch-Verhalten, abgekürzt S/N, ist das Verhältnis zwischen Signal und Rauschen. Oft werden alle Peaks unter einem bestimmten Schwellwert bzw. S/N verworfen, da keine genaue Aussage getroffen werden kann ob es sich um ein falsches Signal handelt oder nicht.

2.3.3 Massenauflösung

Aufgrund der unterschiedlichen Bauweisen der Massenspektrometer gibt es unterschiedliche Verfahren die Auflösung zu berechnen. Grundsätzlich unterscheidet man zwischen der Auflösung von zwei nah beieinander liegenden Peaks und von einem einzelnen Peak. Abbildung 2-13 stellt die zwei Definitionen schematisch dar. Im Allgemeinen wird die Auflösung durch Formel 2.17 angegeben:

$$R = \frac{(m/z)}{\Delta(m/z)} \quad (2.17)$$

Bei der *10 % -Tal Regel* werden zwei nebeneinander liegende, gleich hohe Peaks betrachtet. Die Peak-Kurven überlappen bei 5 % ihrer maximalen Höhe und bilden so ein Tal bei 10 % der maximalen Höhe. Die Auflösung nach dieser Definition berechnet sich demnach aus

$$R_{10\%} = \frac{(m/z)_1}{(m/z)_2 - (m/z)_1} \quad (2.18)$$

Wenn $(m/z)_1 = 1.000$ und $(m/z)_2 = 1.001$ beträgt die Auflösung $R = 1.000$. [5, 22]

Diese Definition ist sinnvoll, wenn die Peak-Form einer Gauss-Kurve entspricht. Als Alternative wird, zum Beispiel bei FT-ICR oder Orbitrap Massenspektrometern, die *Full Width at Half Maximum-Definition*, abgekürzt FWHM angewandt. Das bedeutet es wird

die Breite des Peaks auf halber Höhe gemessen und als $\Delta m/z$ verwendet. Die Auflösung errechnet sich dann mit

$$R_{FWHM} = \frac{(m/z)}{FWHM}. \quad (2.19)$$

Ein Peak bei $m/z = 1.000$ und $FWHM = 0,5$ besitzt eine Auflösung von $R = 2.000$. **[19, 22]** Bei einigen Massenspektrometern, wie FT-ICR oder Orbitrap sinkt die Auflösung mit steigenden m/z – Werten. Es muss dann zusätzlich der m/z – Wert, bei dem die Auflösung gemessen oder berechnet wurde, angegeben werden.

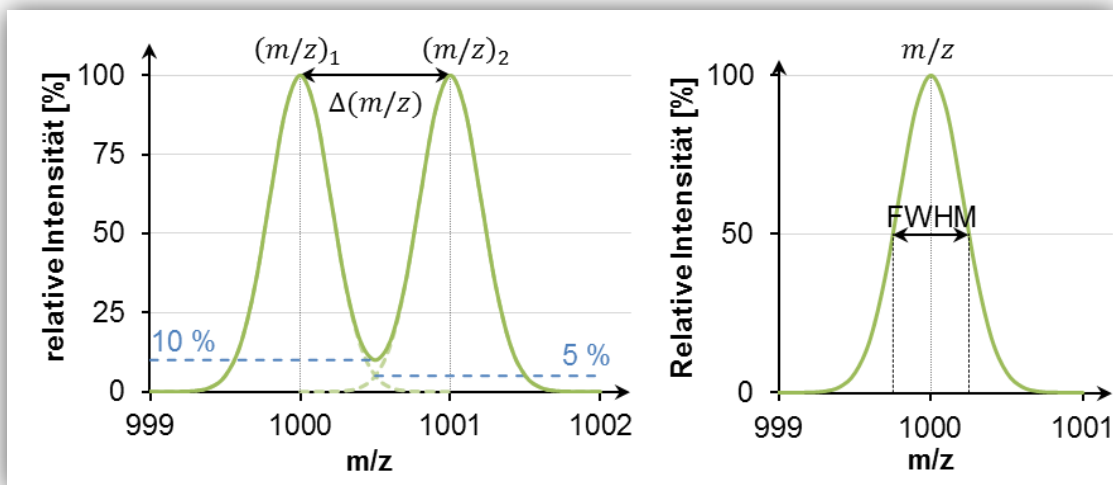


Abbildung 2-13: Massenauflösung mit der 10% - Tal Definition und FWHM Definition

Die Auflösung wird allgemein als $R = (m/z)/\Delta(m/z)$ angegeben. **Links:** Auflösung eines Peaks nach der 10 % - Tal-Regel. Die zwei Peaks überschneiden sich bei 5 % ihrer maximalen Höhe und bilden ein Tal von 10 % der Maximalhöhe. $\Delta(m/z)$ ergibt sich aus der Differenz der m/z – Werte der beiden Peaks $(m/z)_2 - (m/z)_1$. **Rechts:** Auflösung eines Peaks nach der Full Width at Half Maximum Definition. Dabei wird $\Delta(m/z)$ aus der Breite des Peaks auf halber Höhe (FWHM) bestimmt.

Als eine besondere Auflösung wird oft die Einheitsauflösung genannt. Das bedeutet ein Massenspektrometer kann Peaks zwischen zwei aufeinanderfolgenden nominellen Massen auflösen. Also zum Beispiel Peaks bei m/z 50 und m/z 51 oder m/z 1000 und m/z 1001. Oft wird für den Begriff Auflösung auch Auflösungsvermögen verwendet und andersherum. Das Auflösungsvermögen wird auch oft definiert als die Fähigkeit des Massenspektrometers zwei nah beieinanderliegende Peaks zu trennen. Die Differenz $\Delta(m/z)$ wird in Atommasseneinheiten u angegeben. Dabei entspricht $1u = 1,6605402 \cdot 10^{-27} kg$ **[23]**. Oft wird anstatt der Atommasseneinheit auch Dalton (Da) verwendet¹³. Welche Auflösungen Massenspektrometer erreichen können ist in Tabelle 2-2 dargestellt.

¹³ $1Da = 1u$

2.3.4 Massengenauigkeit

Die Massengenauigkeit gibt an wie genau ein Massenspektrometer das m/z -Verhältnis eines Moleküls misst. Die Massendifferenz zwischen dem gemessenen m/z -Wert und dem berechneten m/z -Wert wird als *milli mass units*, abgekürzt *mmu*, angegeben. Die Massengenauigkeit wird in *parts per million*, abgekürzt *ppm*, angegeben und berechnet sich aus

$$ppm = 10^6 \cdot \frac{(m/z)_{\text{gemessen}} - (m/z)_{\text{berechnet}}}{(m/z)_{\text{gemessen}}} \quad (2.20)$$

Je kleiner der Quotient, desto näher entspricht der gemessene m/z -Wert an dem berechneten Wert und desto genauer arbeitet das Massenspektrometer. Eine hohe Massengenauigkeit ist vor allem bei hochauflösenden Massenspektrometern von Bedeutung, da sehr nah beieinander liegende Peaks aufgelöst werden. Zudem ist eine hohe Massengenauigkeit wichtig für die Summenformelbestimmung.

Es ist wichtig die Verhältnisse der Massengenauigkeit über den gesamten Massenbereich zu beachten. Zum Beispiel ergibt eine Massendifferenz von $0,001 \text{ u} = 1 \text{ mmu}$ bei m/z 600 eine Massengenauigkeit von $1,7 \text{ ppm}$, bei m/z 200 beträgt sie schon 5 ppm . [24]

2.3.5 Peak - Formen

Je nach Massenspektrometer kann die ideale Peak-Form durch unterschiedliche Funktionen dargestellt werden. Häufig werden dafür Gauss- oder Lorentz - Funktionen verwendet. In Abbildung 2-14 sind diese beiden Funktionen dargestellt. Für die Gauss-Kurve wurde die nicht normalisierte Gauss Funktion

$$G(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.21)$$

verwendet. Bei Anwendung in einem Massenspektrum beschreibt der Parameter μ den m/z – Wert und σ nimmt Einfluss auf die Breite des Peaks nach Formel 2.22.

$$\sigma = \frac{FWHM}{2 \cdot \sqrt{2 \cdot \ln 2}} = \frac{m/z}{\frac{FWHM}{m/z} \cdot 2 \sqrt{2 \cdot \ln 2}} = \frac{m/z}{R \cdot 2 \sqrt{2 \cdot \ln 2}} \quad (2.22)$$

R beschreibt die Auflösung des Peaks und $FWHM$ ist die Breite des Peaks auf halber Höhe.

Die Lorentz – Funktion wie in Formel 2.23 dargestellt, ergibt eine Lorentz – Kurve. Der Parameter A ist für die Höhe des Peaks verantwortlich und τ entspricht dem $FWHM$. Für x_0 wird der m/z -Wert des Peaks eingesetzt.

$$L(x) = \frac{A}{\pi} \cdot \frac{\tau/2}{(x-x_0)^2 + (\tau/2)^2} = \frac{A}{\pi} \cdot \frac{FWHM/2}{(x-m/z)^2 + (FWHM/2)^2} \quad (2.23)$$

Wird für $A = (\pi \cdot FWHM)/2$ verwendet, besitzt die Lorentz-Kurve eine Höhe von eins.

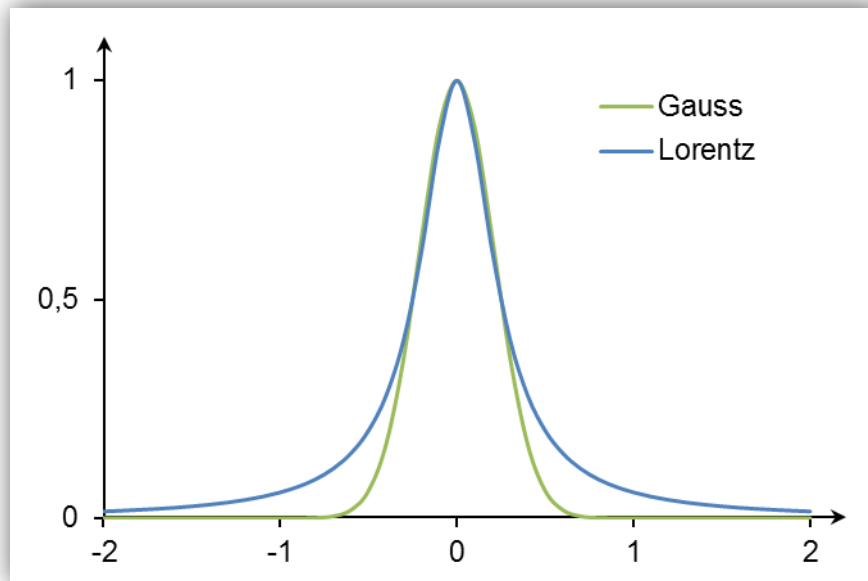


Abbildung 2-14: Peak – Formen nach der Gauss- und Lorentz – Funktion

2.3.6 Zusammenhang zwischen Peak-Abstand und Ladung

Wie im nächsten Kapitel näher erläutert wird besitzt ein Molekül im Massenspektrum immer mehrere Peaks. Der Abstand dieser Peaks zueinander beträgt annähernd ein Dalton, wenn die gemessenen Ionen einfach positiv oder negativ geladen sind. Sind die Ionen mehrfach geladen verringert sich der Abstand um $1/z$ u. Diese Eigenschaft ist vor allem für hochmolekulare Stoffe mit sehr hohen Massen von Vorteil, da sie, wenn sie mehrere Ladungen besitzen, bei geringeren Massen detektiert werden. Das Massenspektrum wird um den Faktor z komprimiert. Man kann den Ladungszustand eines Moleküls anhand des Abstandes zweier benachbarter Peaks die der gleichen Verbindung angehören ermitteln.

3 Simulation von Massenspektren

Liegen von einem gemessenen Massenspektrum Summenformel-Informationen vor, so ist es möglich anhand dieser das Massenspektrum zu simulieren. Dabei wird von jeder Summenformel die Isotopenverteilung berechnet. Vor der Erläuterung der Algorithmen, die dafür zur Verfügung stehen, werden zuerst einige Grundlagen über Isotopie und Massenberechnungen beschrieben.

3.1 Elemente und ihre Isotope

3.1.1 Definition Isotop

Ein Atom besteht grundsätzlich aus drei Komponenten: Elektronen, Protonen und Neutronen. Neutronen und Protonen bilden den Atomkern und Elektronen kreisen in der Elektronenhülle um den Atomkern. Ein Atom eines bestimmten Elements besitzt eine feste Anzahl an Protonen die gleichzeitig der Ordnungszahl Z entspricht. Besitzen Atome die gleiche Anzahl an Protonen aber eine unterschiedliche Anzahl von Neutronen N im Atomkern spricht man von Isotopen. Die Massenzahl $A = Z + N$ gibt nun an wie viele Protonen und Neutronen im Atomkern enthalten sind und unterscheidet somit die Isotope eines Elements. Es wird folgende Schreibweise verwendet: ${}^Z_A\text{ElementSymbol}$. Zum Beispiel bedeutet ${}^{12}_6\text{C}$, dass das Kohlenstoffatom die Ordnungszahl $Z = 6$ und die Massenzahl $A = 12$ besitzt. Diese Schreibweise wird oft zu ${}^Z\text{ElementSymbol}$ abgekürzt, da die Ordnungszahl bei allen Isotopen eines Elements gleich bleibt. Bei schwereren Elementen kann es vorkommen, dass zwei Isotope unterschiedlicher Elemente die gleiche Massenzahl besitzen. [5, 25]

3.1.2 Isotopenhäufigkeiten

Nicht jedes Element besitzt die gleiche Anzahl von Isotopen. Es gibt insgesamt 20 Elemente die nur ein natürlich vorkommendes stabiles Isotop besitzen. Diese nennt man auch *monoisotopische* Elemente. Elemente mit zwei, auch *diisotopische* Elemente genannt, oder mehr Isotopen, auch *polyisotopische Elemente* genannt, besitzen je nach Massenzahl unterschiedliche Massen und kommen unterschiedlich häufig vor. Oft ist das leichteste Isotop eines Elements auch das Häufigste, wie zum Beispiel bei Kohlenstoff, Stickstoff, Sauerstoff oder Wasserstoff. Bei der Angabe der

Isotopenhäufigkeiten unterscheidet man zwischen zwei Normierungsverfahren: Entweder ist das am häufigsten vorkommende Isotop eines Elements auf 100 % normiert oder die Summe der Häufigkeiten aller Isotope eines Elements aufsummiert ergibt 100 %. Man spricht dann von der relativen Häufigkeit. Die Isotopenhäufigkeit gibt an wie oft ein bestimmtes Isotop eines Elements vorkommt. Beide Varianten sind in Tabelle A-1 dargestellt. [5] Die natürlichen Isotopenhäufigkeiten wurden von der *Comission on Isotopic Abundances and Atomic Weights*¹⁴ gemessen und evaluiert. [26] In Abbildung 3-1 sind beispielhaft die relativen Häufigkeiten aller neun Isotope von Xenon gegen ihre Masse aufgetragen.

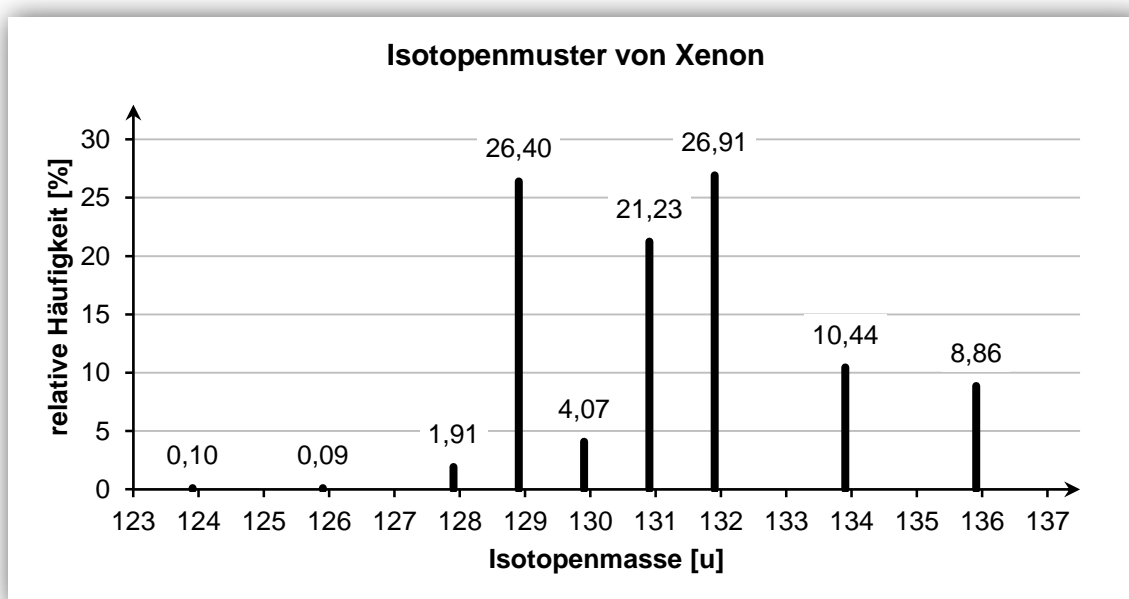


Abbildung 3-1: Isotopenmuster von Xenon

Xenon besitzt insgesamt neun Isotope. Die Isotopenmassen sind in u angegeben. Die genauen relativen Häufigkeiten stehen oberhalb der Linien.

3.2 Isotopenmuster im Massenspektrum

Aufgrund der diisotopsichen oder polyisotopischen Elemente gehören zu einer Verbindung immer mehrere Peaks im Massenspektrum. Grund dafür sind die unterschiedlichen Isotopenmassen und die natürliche Isotopenzusammensetzung der Elemente. Man bezeichnet Verbindungen mit gleicher Elementarzusammensetzung aber unterschiedlicher Isotopenzusammensetzung als *Isotopologe*. [27] Die Abfolge von Peaks in einem Massenspektrum die einer Summenformel zugeordnet werden

¹⁴ Abgekürzt CIAAW, gehört zu IUPAC

können nennt man *Isotopenmuster*. [28] In Abbildung 3-1 wurde bereits das Isotopenmuster von Xenon dargestellt. Die grafische Darstellung ist ähnlich der Darstellung des Massenspektrometers. Es wird die Häufigkeit anstatt der Intensität auf der Ordinate und die Isotopenmasse anstatt des m/z - Wertes auf der Abszisse aufgetragen. Betrachtet man zum Beispiel ein H_2O -Molekül gibt es neun verschiedene Isotopenverteilungen: $^1H_2^{16}O$ | $^1H_2^{17}O$ | $^1H_2^{18}O$ | $^2H_2^{16}O$ | $^2H_2^{17}O$ | $^2H_2^{18}O$ | $^1H\ ^2H^{16}O$ | $^1H\ ^2H^{17}O$ | $^1H\ ^2H^{18}O$. Alle diese Kombinationen besitzen unterschiedliche Massen und Häufigkeiten. Mit welcher Häufigkeit die jeweiligen Isotopenverteilungen auftreten kann durch verschiedene Algorithmen berechnet werden. Diese werden später näher erläutert. Ein Wassermolekül mit $^1H_2^{16}O$ wird sehr viel häufiger vorkommen als $^2H_2^{16}O$. Wo die Isotope in der Verbindung liegen ist bei der Betrachtung der Isotopenmuster uninteressant. Ein Molekül mit $^2H^{16}O\ ^1H$ kommt genauso häufig vor wie ein Molekül mit $^1H^{16}O\ ^2H$. Moleküle deren Isotopenzusammensetzung identisch ist, die Isotope jedoch an unterschiedlichen Positionen im Molekül liegen, nennt man *Isotopomere*. [29]

Aufgrund der Massendifferenz der Isotope von ca. einem Dalton nimmt das Isotopenmuster eines Moleküls einen größeren Massenbereich im Massenspektrum ein. In Abbildung 3-2 ist als Beispiel ein Isotopenmuster von Kaliumbromid und Chloroform dargestellt. Jeder einzelne Peak entspricht einer bestimmten Isotopenverteilung. Die beiden Peaks bei der nominellen Masse von m/z 118 entsprechen den Molekül-Ionen-Peaks von Kaliumbromid und Chloroform. Alle folgenden Peaks des Isotopenmusters entstehen durch die sechs bzw. 15 verschiedenen Isotopenverteilungen von Kaliumbromid und Chloroform. Es sind nicht alle Peaks erkennbar, da einige Isotopenverteilungen sehr geringe Häufigkeiten besitzen. Dies ist auch der Grund warum in einem experimentell gemessenen Massenspektrum nie alle Isotopenpeaks eines Moleküls erkennbar sind. Sie werden durch ihre geringe Signalstärke vom Rauschen überdeckt. Der Molekül-Ionen-Peak wird oft mit M abgekürzt. Der nächste Peak im Isotopenmuster wird dann als $M + 1$ bezeichnet, der übernächste mit $M + 2$ usw. Sind Elemente im Molekül enthalten deren häufigstes Element nicht das leichteste ist, so treten Isotopenpeaks mit kleinerer Masse als die des Molekül-Ionenpeaks auf. Diese werden mit $M - 1, M - 2, \dots$ bezeichnet. Je größer die Moleküle desto häufiger kommt es vor, dass bei einer nominellen Masse zwei oder mehr Isotopenpeaks eines Moleküls entstehen. Zum Beispiel bei m/z 120 die Isotopenverteilungen $^1H^{12}C^{35}Cl_2^{37}Cl$ und $^2H^{13}C^{35}Cl_3$ mit m/z

119,9114 und m/z 119,924. Zwei Moleküle werden jedoch nie die gleiche Masse besitzen aufgrund der verschiedenen Massendefekte der Isotope.

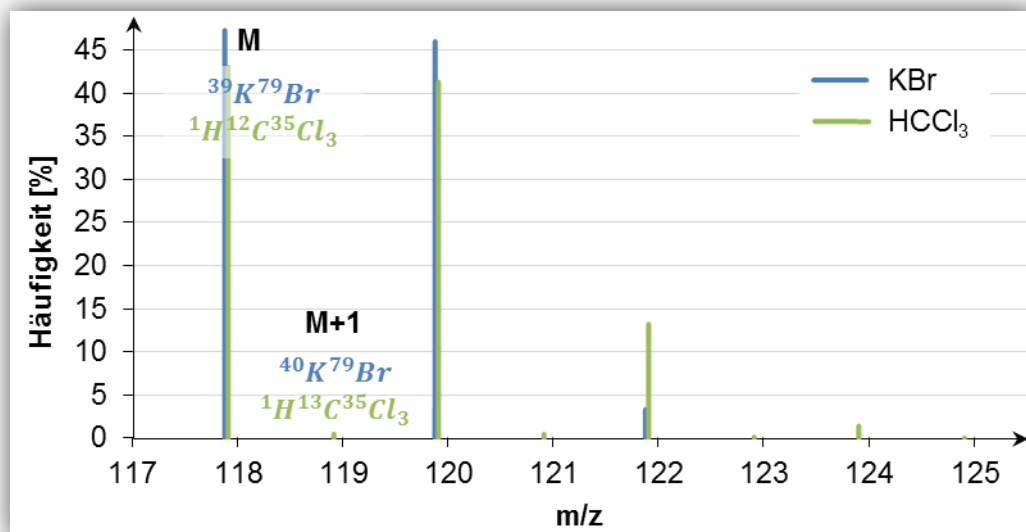


Abbildung 3-2: Isotopenmuster von Chloroform und Kaliumbromid

Es sind die theoretischen Isotopenmuster von Kaliumbromid (blau) und Chloroform (grün) als einfach positiv geladene Ionen dargestellt. Die Molekül-Ionen-Peaks (M) der beiden Moleküle besitzen dieselbe nominelle Masse von 117 u. Das Isotopenmuster von Kaliumbromid besteht aus sechs Peaks und erstreckt sich über einen Massebereich von 117 bis 122 u und das von Chloroform aus 15 Peaks mit einem Massebereich von 117 bis 126 u. Aufgrund sehr geringer Häufigkeiten von weniger als $1 \cdot 10^{-5} \%$ sind nicht alle Peaks erkennbar.

Ob alle Peaks bei einer nominellen Masse von einem Massenspektrum als einzelne Peaks dargestellt werden können hängt von der Auflösung und Massengenauigkeit des Geräts ab. Ist die Auflösung nicht ausreichend addieren sie die jeweiligen Anteile der Häufigkeiten der Peaks und es entstehen aggregierte Peaks deren m/z -Wert dem Mittel aller beteiligten Isotopenverteilungen entspricht.

3.3 Berechnung von Molekül- und Ionenmassen

Es gibt drei verschiedene Möglichkeiten die Masse eines Atoms anzugeben: die nominelle, relative und exakte Isotopenmasse. Daraus ergeben sich für jedes Molekül auch drei verschiedene Massenangaben die in Tabelle 3-1 ist am Beispiel CO_2 berechnet worden. Als Einheit wird die *unified atomic mass*, abgekürzt *u*, verwendet. Nach der IUPAC Definition ist 1 *u* definiert als $1/12$ der Masse des ^{12}C Atoms und entspricht $1,660538921 \cdot 10^{-27} \text{ kg}$. Für die *unified atomic mass* wird oft auch die Einheit *Dalton*, abgekürzt *Da*, verwendet. [10]

3.3.1 Nominelle Masse

Die *nominelle Masse* $m_N \in \mathbb{N}$ eines Elements ergibt sich aus der gerundeten Masse des häufigsten natürlich vorkommenden stabilen Isotops. Der numerische Wert der nominellen Masse stimmt mit dem der Massenzahl überein. Die nominelle Masse eines Moleküls ergibt sich aus der Addition aller nominellen Massen der enthaltenen Elemente:

$$m_{N,Molekül} = \sum_{i=1}^k (c_i \cdot m_{N,i}) \quad \text{mit} \quad m_{N,i} = [m_{I,i}] \quad (3.1)$$

wobei c_i die Anzahl und $m_{N,i}$ die nominelle Masse des Elements i im Molekül, k die Anzahl der Elemente in der Verbindung und $m_{I,i}$ die Masse des häufigsten Isotops des Elements i darstellt. [10, 5]

3.3.2 Relative Masse

Die Massen der Elemente und deren Isotope sind jedoch keine ganzzahligen Werte. Eine genauere Berechnung der Masse eines Atoms wird durch die Verwendung der *relativen Masse* ¹⁵ $m_R \in \mathbb{R}^+$ erreicht. Sie ergibt sich aus dem Mittelwert der Isotopenmassen eines Elements gewichtet nach ihren Isotopenhäufigkeiten:

$$m_{R,Molekül} = \sum_{i=1}^k (c_i \cdot m_{R,i}) \quad \text{mit} \quad m_{R,i} = \frac{\sum_{j=1}^l (a_j \cdot m_{I,j})}{\sum_{j=1}^l a_j} \quad (3.2)$$

Wobei c_i die Anzahl und $m_{R,i}$ die relative Masse des Elements i im Molekül, a_j die Isotopenhäufigkeit und $m_{I,j}$ die Isotopenmasse des Isotops j sowie l die Anzahl der Isotope des Elements darstellt. Die relative Masse wird verwendet, wenn keine Information über die Isotopenzusammensetzung bekannt ist. Da sich die Massen einzelner Isotope des gleichen Elements jedoch um ≥ 1 Da unterscheiden, kann es bei größeren Molekülen vorkommen, dass die relative Masse um einige Dalton von der exakten Masse abweichen. [16, 5, 9]

¹⁵ Engl.: average mass

3.3.3 Exakte Masse

Ist in einem Molekül die genaue Isotopenzusammensetzung bekannt oder will man die genaue monoisotopische Masse¹⁶ berechnen werden die jeweiligen Isotopenmassen verwendet. Die dadurch ermittelte *exakte Masse*¹⁷ $m_E \in \mathbb{R}^+$ ist die genaueste Angabe einer Molekülmasse und wird wie folgt berechnet:

$$m_{E,Molekül} = \sum_{i=1}^k (c_i \cdot m_{I,i}) \quad (3.3)$$

Wobei $m_{I,i}$ die exakte Masse des Isotops i und c_i die Anzahl des Isotops i im Molekül. Die exakten Massen der Isotope sind relativ zum ^{12}C -Isotop, welche eine definierte Masse von 12,000 u besitzt, gemessen [5, 16, 9] In Tabelle 3-1 sind am Beispiel von CO_2 die Massen nach den unterschiedlichen Massendefinitionen berechnet worden.

Tabelle 3-1: Beispiel der unterschiedlichen Massendefinitionen

Die Tabelle zeigt am Beispiel von CO_2 die Werte der nominellen, mittleren und Isotopenmassen. Isotopenmassen wurden aus [5] und mittlere Masse aus [30] entnommen.

Isotop/Molekül/Ion	nominelle Masse [u]	mittlere Masse [u]	Isotopenmasse [u]
^{12}C	12	12,0108	12
^{13}C			13,003355
^{16}O	16	15,994	15,994915
^{17}O			16,999132
^{18}O			17,999116
$^{12}\text{C}^{16}\text{O}_2$	44	43,98983	44,0096
$^{13}\text{C}^{17}\text{O}_2$	44	43,98983	49,001587

¹⁶ Bei der monoisotopischen Masse eines Moleküls werden jeweils die Massen des häufigsten Isotops der Elemente addiert.

¹⁷ Engl.: exact mass

3.3.4 Masse-zu-Ladungsverhältnis

Da bei der Massenspektrometrie nur das m/z - Verhältnis geladener Teilchen gemessen werden kann, ist es wichtig auch die Masse der fehlenden bzw. zusätzlichen Elektronen einzurechnen. Die Masse eines Elektrons beträgt $m_e = 0,000548 \text{ u}$. Vor allem bei FT-ICR- oder Orbitrap- Massenspektrometern mit Massengenauigkeiten kleiner als 10^{-3} u ist es wichtig die Massendifferenz die durch eine Ionisierung entsteht mit einzurechnen. Bei positiven Ionen wird das Produkt aus Anzahl der Ladungen z und Elektronenmasse m_e subtrahiert, bei negativer Ladung addiert. [5, 31] Die Masse $m_{E,Ion}$ wird zunächst mit

$$m_{E,Ion} = \sum_{i=1}^k (c_i \cdot m_{I,i}) \pm z \cdot m_e \quad (3.4)$$

berechnet, wobei c_i die Anzahl und $m_{I,i}$ die Isotopenmasse des Isotops i darstellt. Um das m/z Verhältnis zu erhalten wird die Ionenmasse anschließend durch die Anzahl der Ladungen dividiert:

$$m/z = \frac{m_{E,Ion}}{z} \quad (3.5)$$

3.3.5 Der Massendefekt

Addiert man die Massen der Nukleonen¹⁸ eines Atoms mit bestimmter Massenzahl und vergleicht das Ergebnis mit der exakten Masse so ist die exakte Masse stets niedriger. Diese Differenz zwischen exakter Masse und der Summe aller beteiligten Nukleonen nennt man Massendefekt¹⁹. Dieser Massenverlust ist äquivalent zur Bindungsenergie, die aufgebracht werden muss, um aus den Neutronen und sich gegenseitig abstoßenden Protonen einen Atomkern zu bilden. Jedes Isotop besitzt einen einzigartigen Massendefekt. Durch diese Eigenschaft ist es bei der hochauflösenden Massenspektrometrie möglich Verbindungen mit gleicher nomineller Masse aber unterschiedlicher exakter Masse zu detektieren und zu unterscheiden. [9, 32]

In der Massenspektrometrie weicht man von der physikalischen Definition ab und definiert den Massendefekt als Differenz zwischen der nominellen und der exakten Masse eines Atoms.

¹⁸ Als Nukleon bezeichnet man die Komponenten die den Atomkern, also Protonen und Neutronen, bilden.

¹⁹ In einigen Literaturquellen werden auch Elektronen mit in die Definition des Massendefekts einbezogen. [6]

$$m_D = m_N - m_E \quad (3.6)$$

Es treten sowohl positive als auch negative Massendefekte auf, mit Ausnahme von ^{12}C welches durch seine definierte Masse von 12 u einen Massendefekt von 0 u besitzt. In Abbildung 3-3 ist der Massendefekt in Abhängigkeit von der nominellen Masse dargestellt. [8, 33]

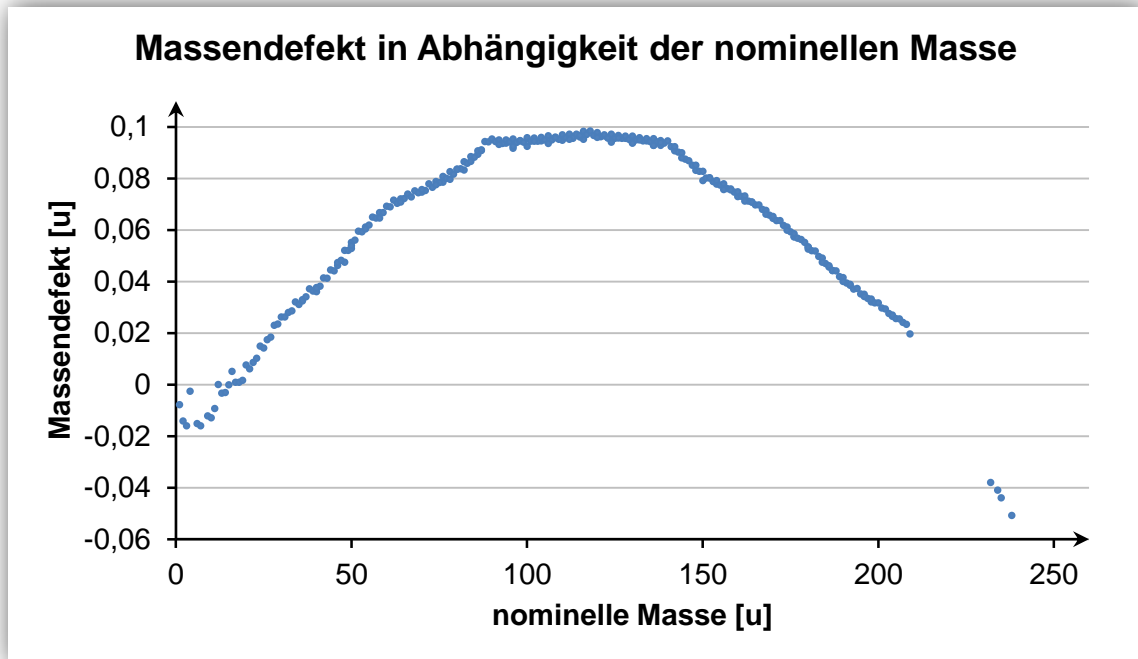


Abbildung 3-3: Massendefekt pro Nukleon aller Isotope

Der Massendefekt wie er in der Massenspektrometrie benutzt wird, berechnet sich aus der Differenz zwischen nomineller und exakter Masse eines Isotops. In der Abbildung ist der Massendefekt aller stabilen Isotope dargestellt.

3.4 Algorithmen zur Isotopenmusterberechnung

Es gibt verschiedene Methoden das Isotopenmuster eines Moleküls zu berechnen. Im Folgenden werden einige dieser Methoden beschrieben und algorithmisch dargestellt. Sie beziehen sich auf die Berechnung der Isotopenverteilungen und deren relativen Häufigkeiten. Voraussetzung für diese Berechnungen ist eine gegebene Summenformel. Um ein simuliertes Massenspektrum darzustellen benötigt man zusätzlich die exakte Masse der jeweiligen Verbindungen.

3.4.1 Binomiale und Polynomiale Methode

Für die Berechnung eines Isotopenmusters ist es notwendig, die Häufigkeiten für jede mögliche Isotopenverteilung zu berechnen. Besteht ein Molekül aus n Atomen eines diisotopischen Elements A , wie zum Beispiel C_{60} , sind die relativen Häufigkeiten der Isotopenverteilungen durch

$$(a_1 + a_2)^n \quad (3.7)$$

gegeben. Dabei sind a_1 und a_2 die relativen Häufigkeiten der Isotope des Elements A . Ausmultipliziert ergibt der Term $n + 1$ Summanden. Für ein Wasserstoff-Molekül H_2 entstehen die vier Summanden

$$\left(h(^1H) + h(^2H)\right)^2 = h(^1H)^2 + h(^1H)h(^2H) + h(^2H)h(^1H) + h(^2H)^2, \quad (3.8)$$

wobei $h(\cdot)$ die jeweiligen relativen Häufigkeiten der Wasserstoff-Isotope darstellen. Jeder Summand ergibt die relative Häufigkeit einer bestimmten Isotopenkombination. Dabei wird auch berücksichtigt, an welcher Position die Isotope liegen²⁰. Da Isotopomere die gleiche relative Häufigkeit und Masse besitzen, können diese auch zusammengefasst werden, sodass nur Summanden übrig bleiben, die Isotopologe repräsentieren. Bei Wasserstoff würde also $h(^1H)h(^2H) + h(^2H)h(^1H) = 2 \cdot h(^1H)h(^2H)$ ergeben. Die Verwendung dieser binomialen Gleichung ist nur für diisotopische Elemente geeignet. Durch Verallgemeinerung und Verwendung polynomialer Gleichungen ist es möglich auch polyisotopische Elemente mit einzubeziehen. Besteht ein Molekül aus mehr als einem Element müssen die einzelnen polynomialen Terme multipliziert werden:

²⁰ Oft werden die Summanden, die zusammengefasst werden können auch als „like-terms“ beschreiben.

$$(a_1 + a_2 + \dots + a_k)^{nA} \cdot (b_1 + b_2 + \dots + b_k)^{nB} \cdot (c_1 + c_2 + \dots + c_k)^{nC} \cdot \dots \quad (3.9)$$

Die Variablen $a_i, b_i, c_i, i = 1, \dots, k$ stellen dabei die relativen Häufigkeiten der Isotope der jeweiligen im Molekül enthaltenen Elemente A, B, C und nA, nB, nC die Anzahl der Atome jedes Elements dar.

Der Nachteil dieser Methode besteht darin, dass bei großen Molekülen der Rechenaufwand sehr groß wird, da die Anzahl der Summanden für jedes Element exponentiell wächst. Zum Beispiel besitzt ein Molekül bestehend aus $C_n H_m O_o$ insgesamt $2^n 2^m 3^o$ mögliche Isotopenkombinationen. Das anschließende Zusammenfassen der Isotopomere kostet zusätzlichen Aufwand und Rechenzeit. Daher ist es sinnvoll diese Methode nur bei kleinen Molekülen anzuwenden. [34, 35]

3.4.2 Multinomiale Methode

Einen polynomialen Term $(a_1 + a_2 + \dots)^{nA}$ kann man mit Hilfe der multinomialen Verteilung als Summe über alle Isotopenverteilungen p

$$(a_1 + a_2 + \dots + a_k)^{nA} = \sum_{i=1}^p \left(\frac{nA!}{n_{i,1}! \cdot n_{i,2}! \cdot \dots \cdot n_{i,k}!} \cdot (a_1)^{n_{i,1}} (a_2)^{n_{i,2}} \dots (a_k)^{n_{i,k}} \right) \quad (3.10)$$

darstellen. Dabei stellen $a_1 \dots a_k$ die relativen Häufigkeiten der Isotope des Elements A , n die Anzahl der Atome des Elements A und $n_{i,1} \dots n_{i,k}, \sum_{j=1}^k n_{i,j} = n$, die Anzahl der Isotope in einer Isotopenverteilung dar. Mit dieser Formel lassen sich die Häufigkeiten der Isotopenverteilungen direkt berechnen. Jeder Summand repräsentiert die Häufigkeit einer Isotopenverteilung. Diese Berechnung wurde 1983 von James A. Yergey beschrieben. [35, 34]

Der Vorteil dieser Methode ist, dass für jedes Element in einer Summenformel die Häufigkeiten getrennt berechnen lassen und der Aufwand der Berechnung gegenüber der polynomialen Methode sehr viel geringer ist. Problematisch ist jedoch die Berechnung der Fakultät bei Molekülen mit großer Anzahl an Atomen pro Element. Um die Berechnung der Fakultät großer Zahlen zu reduzieren, werden die Häufigkeiten zweier Isotopenverteilungen eines Elements $n_{i,1}, \dots, n_{i,k}$ und $n_{l,1}, \dots, n_{l,k}$ dividiert und man erhält:

$$H_{n_i} = H_{n_l} \cdot \frac{n_{l,1}! \cdot n_{l,2}! \cdot \dots \cdot n_{l,k}!}{n_{i,1}! \cdot n_{i,2}! \cdot \dots \cdot n_{i,k}!} \cdot (a_1)^{n_{i,1}-n_{l,1}} (a_2)^{n_{i,2}-n_{l,2}} \dots (a_k)^{n_{i,k}-n_{l,k}} \quad (3.11)$$

Dabei sind H_{n_i} und H_{n_l} die Häufigkeiten der zwei Isotopenverteilungen und $a_1 \dots a_k$ die relativen Häufigkeiten der Isotope des Elements A . Zum Beispiel betrachte man die

zwei Verteilungen $n_{i,1}, n_{i,2}, n_{i,3} = 47, 3, 0$ und $n_{l,1}, n_{l,2}, n_{l,3} = 47, 2, 1$ eines Elements so erhält man als Ergebnis für die Häufigkeit H_n

$$H_n = H_m \cdot \frac{47! \cdot 3! \cdot 0!}{47! \cdot 2! \cdot 1!} \cdot (a_1)^{47-47} (a_2)^{3-2} (a_3)^{0-1} = 3H_m (a_2)^1 (a_3)^{-1}. \quad (3.12)$$

Voraussetzung für eine solche Rechnung ist ein gegebenes H_m . Zu Beginn der Berechnungen der Häufigkeiten der Isotopenverteilungen für ein Element ist es daher sinnvoll die Häufigkeit der monoisotopischen Verteilung, wie zum Beispiel $^{16}\text{O}_{50}$ zu berechnen. Anhand dieser können alle anderen Häufigkeiten nach Formel 3.11 berechnet werden. [35]

Wie schon erwähnt können in einem experimentell gemessenen Massenspektrum nicht alle Isotopenpeaks aufgelöst werden, da die entsprechenden Isotopenverteilungen nicht häufig genug vorkommen und das Signal nicht von Rauschen unterschieden werden kann. Unter dieser Annahme kann man die Berechnung der Isotopenmuster nach der Methode nach Yergey weiter vereinfachen, indem man bestimmte Isotopenverteilungen von der Berechnung ausschließt wenn die Häufigkeit unter einen bestimmten Schwellwert sinkt. Dies wird auch als Pruning bezeichnet. Betrachtet man nicht die einzelnen sondern aggregierte Isotopenpeaks, kann diese Reduktion jedoch dazu führen dass die resultierende Höhe der Peaks bei einer nominellen Masse niedriger ausfällt. Bedingt wird dies durch die fehlenden, wenn auch kleinen Anteile weniger häufig vorkommender Isotopenverteilungen.

Die multinomiale Methode der Isotopenmusterberechnung produziert ein Linienspektrum in dem fast alle, abhängig vom gewählten Schwellwert, theoretischen Isotopenpeaks berechnet werden können. Für eine Darstellung als Profilspektrum muss eine Peak-Shape Funktion auf die berechneten Peaks angewandt werden.

3.4.3 Fourier-Transformation basierte Berechnungen

Ähnlich wie die Umwandlung der Zeit-Domäne in die Frequenz-Domäne bei der FT-ICR kann man die Fourier-Transformation zur Berechnung von Isotopenmustern nutzen. Dieser Algorithmus geht auf die Forschungsgruppe von Alan L. Rockwood zurück und wurde 1995 erstmalig vorgestellt. [36] In Formel 3.10 werden ausschließlich Informationen über Isotopenhäufigkeiten in die Berechnung der Isotopenverteilung einbezogen. Rockwood erweitert diesen Term indem er einen symbolischen Identifikator I_x für jedes Isotop $a_1, a_2, \dots, b_1, b_2, \dots$ der Elemente A, B, \dots einsetzt:

$$(a_1 I_{a_1} + a_2 I_{a_2} + \dots + a_k I_{a_k})^n \cdot (b_1 I_{b_1} + b_2 I_{b_2} + \dots + b_k I_{b_k})^o \cdot \dots \quad (3.13)$$

Diese zusätzliche Information ermöglicht es, die Masse der einzelnen Isotopen(-verteilungen) während der Berechnung der Häufigkeiten mit einzubeziehen. Für die Anwendung einer Fourier Transformation substituiert man diesen Identifier mit einer Funktion $e^{i2\pi m_x \mu}$ bei der m_x die Masse des jeweiligen Isotops darstellt und $i = \sqrt{-1}$ die imaginäre Zahl. Nach Anwendung der multinomialen Methode führt diese Substitution zu einer Funktion die abhängig vom Parameter μ ist:

$$f(\mu) = \sum \left(\frac{nA!}{nA_1! nA_2! \dots nA_k!} (a_1)^{nA_1} (a_2)^{nA_2} \dots (a_k)^{nA_k} \cdot e^{i2\pi(m_{a_1} nA_1 + m_{a_2} nA_2 + \dots + m_{a_k} nA_k) \mu} \right) \cdot \sum \left(\frac{nB!}{nB_1! nB_2! \dots nB_k!} (b_1)^{nB_1} (b_2)^{nB_2} \dots (b_k)^{nB_k} \cdot e^{i2\pi(m_{b_1} nB_1 + m_{b_2} nB_2 + \dots + m_{b_k} nB_k) \mu} \right) \cdot \dots \quad (3.14)$$

Die Summation erfolgt wieder über alle Isotopenverteilungen p . Der Index wurde aus Gründen der besseren Übersicht weggelassen. Die Summanden beschreiben die Häufigkeit einer Isotopenverteilung des Moleküls und haben folgende Form:

$$\left(\frac{nA!}{nA_1! nA_2! \dots nA_k!} (a_1)^{nA_1} (a_2)^{nA_2} \dots (a_k)^{nA_k} \right) \left(\frac{nB!}{nB_1! nB_2! \dots nB_k!} (b_1)^{nB_1} (b_2)^{nB_2} \dots (b_k)^{nB_k} \right) \cdot \dots \cdot e^{i2\pi(m_{a_1} nA_1 + \dots + m_{a_k} nA_k + m_{b_1} nB_1 + \dots + m_{b_k} nB_k) \mu} \quad (3.15)$$

Aufgrund der Periodizität dieser Funktion, ist eine ähnliche Ausgangssituation wie zum Beispiel bei der FT-ICR MS gegeben. Durch eine anschließende Fourier-Transformation in die Massen-Domäne wird die periodische Funktion in einen Peak bei der Masse $m_{a_1} nA_1 + \dots + m_{a_k} nA_k + m_{b_1} nB_1 + \dots + m_{b_k} nB_k$ umgewandelt. [36]

Der Parameter μ ist ein Zahlen-Array, welches die Anzahl der Datenpunkte bestimmt und die Auflösung des simulierten Spektrums beeinflusst. Man spricht auch von der μ -Domäne [34] Die Verteilung jedes einzelnen Elements aus Formel 3.15 lässt sich nun aufgrund des Konvolutions-Theorems zusammenfassen zu:

$$f_{A nA B nB \dots}(\mu) = [f_A(\mu)]^{nA} [f_B(\mu)]^{nB} \dots \quad (3.16)$$

Es besagt, dass die Konvolution in einer Domäne (hier μ) mit der Multiplikation in der anderen (hier m) korrespondiert.

Um eine bestimmte Peakform in die Berechnung mit einzubeziehen wird die entsprechende Peak-Funktion $S(m)$, wie die Lorentz- oder Gaußskurve in die μ -Domäne durch *inverse Fourier Transformation (IFT)* übersetzt, mit Formel 3.16 multipliziert und anschließend die Transformation zurück in den Massen-Domäne durchgeführt. Dieser Prozess ist in Formel 3.17 angedeutet.

$$s(\mu) = IFT[S(m)] \rightarrow F(m) = FT[s(\mu)f(\mu)] \quad (3.17)$$

Für eine möglichst schnelle Berechnung der Isotopenverteilungen und deren Häufigkeiten wird in der Praxis die schnelle Fourier-Transformation verwendet. Als Voraussetzung dafür muss das Array μ aus $n^2, n \in \mathbb{N}$ Datenpunkten bestehen. Als weitere Vereinfachung wird ein so genanntes *heterodyning* auf die Funktion $f(\mu)$ angewendet. Dabei wird die Masse der Isotopenverteilung in die Nähe von null verschoben:

$$F(m - m_0) = FT[e^{-m_0(i2\pi)\mu} s(\mu)f(\mu)] \quad (3.18)$$

Für m_0 wird zum Beispiel die mittlere Masse des Moleküls, für die die Isotopenverteilung berechnet wird, eingesetzt. Nach der Fourier-Transformation in die Massen.-Domäne wird m_0 wieder aufaddiert. Eine bessere Performance bei der Berechnung kann man ebenfalls durch die Verwendung von Polarkoordinaten anstatt kartesischer Koordinaten erreichen, da im komplexen Raum gerechnet werden kann.

Der Parameter μ bestimmt die Anzahl der Datenpunkte und wird bestimmt durch das Produkt aus Massenbereich und Anzahl benötigter Datenpunkte pro Dalton. Der Massenbereich sollte dabei so gewählt werden, dass alle Peaks des Isotopenmusters innerhalb dieses Bereichs liegen. Wird der Massenbereich zu klein gewählt so werden die Peaks außerhalb dieses Bereichs, bedingt durch die Eigenschaften der Fourier-Transformation, in das Isotopenmuster an einer falschen Stelle rückgefaltet. Formel 3.19 gibt eine Möglichkeit der Berechnung des Massenbereichs wieder.

$$\text{Massenbereich} = k(1 + \sigma^2)^{1/2} \quad (3.19)$$

Dabei ist σ die Standardabweichung des Moleküls gegeben durch die Addition der quadratischen Mittel der jeweiligen Isotopenverteilungen. Der Faktor k ist ein Skalierungsfaktor, welcher abhängig von der Größe des Moleküls gewählt wird. Für größere Moleküle wird oft $k = 10$ gesetzt. Bedingt durch die Verwendung der FFT sollte die resultierende Anzahl an Datenpunkten zu einem Wert n^2 , wie zum Beispiel 1024 aufgerundet werden. [34, 37]

Mit dieser Methode können Profilspektren simuliert werden, die durch ihre Peak-Form und variierbare Peak-Breiten an die Auflösung experimenteller Spektren anpassbar sind und dadurch ein besserer Vergleich zwischen Theorie und gemessenem Spektrum möglich ist. Sie wurde von Rockwood et. al. In dem Programm Mercury implementiert. [38, 37] Nachteil dieses Programm ist es, dass nur aggregierte

Profilspektren generiert werden. Es ist zwar theoretisch möglich die Feinstruktur, d.h. die Isotopenpeaks aufzulösen, jedoch benötigt dies viel mehr Rechenzeit und Speicherkapazität. Eine Lösung dieses Problems ist es, nur einen bestimmten Massenbereich wie z.B. 0.2 Da des Spektrums zu betrachten und innerhalb dieses Bereichs die Peaks mit hoher Auflösung zu berechnen. Dies bedeutet, dass das Spektrum durch viele einzelne Berechnungen - für jede nominelle Masse eine - erstellt wird und führt zu langen Rechenzeiten. [38]

Eine andere Möglichkeit die Isotopenfeinstruktur des kompletten Spektrums aufzulösen kann durch Komprimierung des Massenbereichs erreicht werden. Betrachtet man experimentelle Massenspektren, erkennt man, dass alle Isotopenpeaks immer nahe der nominellen Masse gruppiert sind. Zwischen zwei nominellen Massen existiert ein Bereich von beispielsweise $r = 0,8 \text{ Da}$ in dem keine Isotopenpeaks auftreten. Diese Leerräume werden mit

$$m^* = m - n_m \cdot r \quad (3.20)$$

von der Masse m eines Moleküls subtrahiert. Der Faktor n_m berechnet sich aus der Differenz der nominellen Masse des monoisotopischen Peaks und der nominellen Masse des zu komprimierenden Peaks. In Abbildung 3-4 ist am Beispiel von Kaliumbromid und Chloroform die Komprimierung dargestellt. Mit Hilfe von Formel 3.20 verkürzt sich der Massenbereich des Spektrum von $117.88 \text{ Da} - 125.91 \text{ Da}$ auf $117.88 \text{ Da} - 119.51 \text{ Da}$. Der monoisotopische Peak verbleibt an der gleichen Stelle im Spektrum. Der neue kleinere Massenbereich kann durch die Multiplikation von Formel 3.19 mit $(1 - r)$ ermittelt werden. [39]

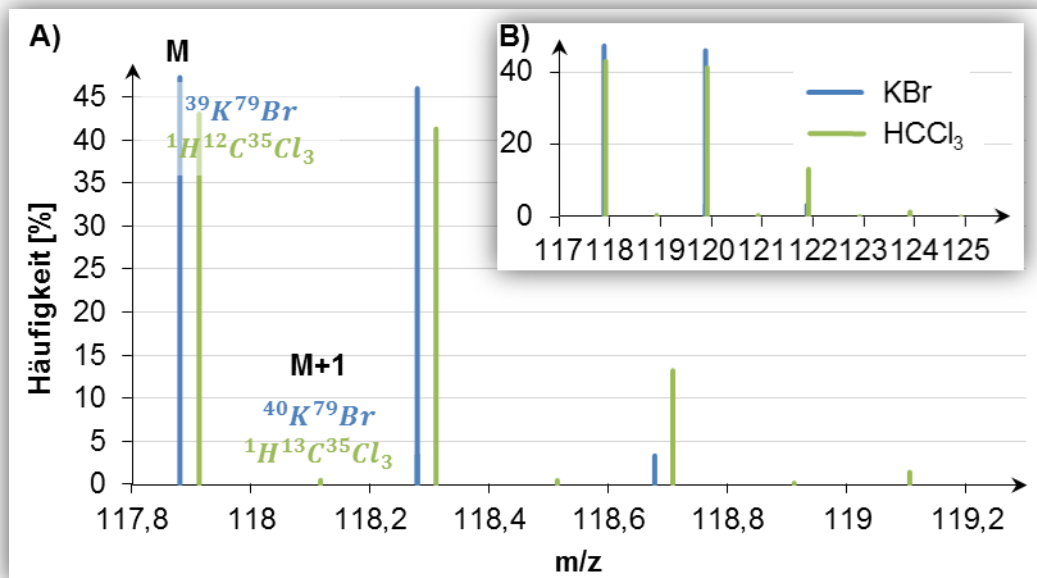


Abbildung 3-4: Komprimierung des Massenspektrums

Durch das Beseitigen von Leerraum wird das Massenspektrum auf einen kleineren Massenbereich komprimiert. In diesem Beispiel wurde $r = 0.8$ gewählt. A) Das komprimierte Massenspektrum von Kaliumbromid und Chloroform. B) Das Originalspektrum der beiden Moleküle. Der Molekülpeak M befindet sich an der gleichen Position. Alle folgenden Isotopenpeaks wurden nach links verschoben. Die M+n Peaks sind um $n \cdot 0.8 \text{ Da}$ nach links verschoben worden.

3.4.4 Zweidimensionale Fourier-Transformation

Die im vorigen Kapitel vorgestellte auf Fourier-Transformation basierte Berechnung der Isotopenmuster und deren Verbesserungen sind nur von einem Parameter m bzw. μ , je nach Domäne, abhängig. Das Problem dieser Methoden ist jedoch nach wie vor die effiziente und schnelle Berechnung bei sehr hoher Auflösung. [39]

Ein Spektrum kann man auch in zwei Komponenten betrachten: Die erste beschreibt die grobe Struktur des Massenspektrums in Abständen von $\sim 1 \text{ Da}$. Die zweite Komponente ist die Feinstruktur des Spektrums im Milli-Dalton Bereich. Diese zweidimensionale Betrachtung ermöglicht es die Berechnung der Isotopenverteilung eines Moleküls mit der Summenformel $A_{nA}B_{nB} \dots Z_{nZ}$ in zwei Dimensionen aufzuteilen. In der μ Domäne ergibt sich daraus die Formel

$$f_{A_{nA}B_{nB} \dots Z_{nZ}}(u, w) = [f_A(u, w)]^{nA} [f_B(v, w)]^{nB} \dots [f_Z(v, w)]^{nZ} \quad (3.21)$$

mit

$$f_s(u, w) = \sum_{s \in \text{Isotop}} h_{n_s, \Delta m_s} e^{-2\pi i (n_s \cdot u + \Delta m_s \cdot w)}. \quad (3.22)$$

Wobei h_{n_s} die Häufigkeit der jeweiligen Isotope, $\Delta m_s = n_s - m_s$ der Massendefekt des Isotops s , n_s die nominelle Masse des Isotops s und $i = \sqrt{-1}$. Durch eine Fourier-Transformation in die Massendomäne mit den Parametern n und dem Massendefekt $\Delta m = m - n$ ergibt sich eine 2D Isotopenverteilung von

$$F(n, \Delta m) = \sum_{s \in \{A, B, \dots, Z\}} h_{n_s} \delta_{n, n_s} \delta(\Delta m - \Delta m_s). \quad (3.23)$$

Die Funktionen δ_{n, n_s} ist das Kronecker Delta²¹ und $\delta(\Delta m - \Delta m_s)$ die Dirac Delta Funktion²². Die Summe aus allen $F(n, \Delta m)$ ergibt das Massenspektrum des Moleküls $A_{n_A} B_{n_B} \dots Z_{n_Z}$.

Wie in Formel 3.17 schon beschrieben, wird hier ebenfalls eine Peak-Form Funktion mit in die Berechnung einbezogen werden. Der allgemeine Ablauf der ursprünglichen Fourier-Methode bleibt unverändert. Zuerst wird die Fourier-Transformation auf alle beteiligten Elemente angewandt, diese anschließend entsprechend der Summenformel und mit der Peak-Form Funktion multipliziert und danach mit Hilfe der inversen FT wieder zurück in die Massendomäne transformiert. Das Ergebnis ist eine 2D Isotopenverteilung wie in Abbildung 3-5 schematisch dargestellt. Für die übliche Darstellung eines Massenspektrums im m/z –Intensitäts-Diagramm müssen die Massen durch Addition der Massendefekte sowie der nominellen Massen wiederhergestellt und die Häufigkeiten, wenn nötig, addiert werden. [40]

²¹ $\delta_{n, n_s} = \begin{cases} 1 & \text{wenn } n = n_s \\ 0 & \text{sonst} \end{cases}$

²² $\delta(\Delta m - \Delta m_s) = 1 \text{ wenn } \Delta m = \Delta m_s$

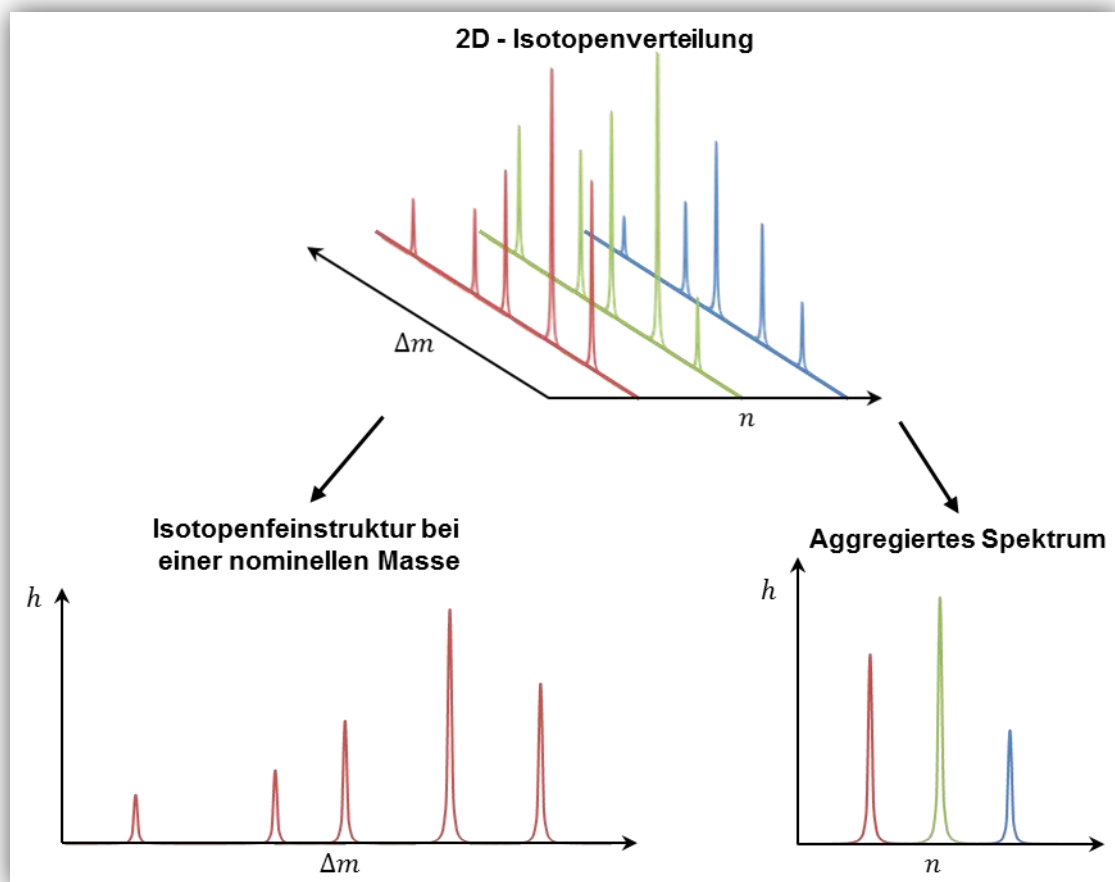


Abbildung 3-5: Schematische Darstellung der 2D – FT Methode

Das Schema verdeutlicht die Vorgehensweise des 2D-FT- Algorithmus. Die Isotopenfeinstruktur (farbige Spektren) und die aggregierten Peaks (einzelne farbige Peaks) an den nominellen Massen werden getrennt betrachtet und durch zwei die zwei Parameter – Massendefekt $\Delta m = n - m$ bzw. nominelle Masse n in der Berechnung repräsentiert.

Durch die Verwendung einer zweidimensionalen FT Methode konnte der ursprüngliche Algorithmus von Rockwood et. al. [38] um das 10 bis 270fache verbessert werden. Beispielsweise benötigt der Algorithmus für Rinder-Insulin $C_{254}H_{378}N_{65}O_{72}S_6$ bei einer Auflösung von $3 \cdot 10^7$ nur 0.484 Sekunden. Die *Mercury Software* von Rockwood et. al. ist mit 3.6 Sekunden um einiges langsamer. [40]

4 Datensatz und Dateiformate

4.1 Spektren Datensatz

4.1.1 Massenspektrometer und Methode

Die zur Verfügung gestellten MS-Daten wurden mit Hilfe eines *Bruker Daltonics Solarix* FT-ICR-Massenspektrometer aufgenommen.

Als Ionisierungsmethode wurde die Elektrospray-Ionisierung gewählt. Es wurde eine Messung ohne und eine mit Fragmentierung durch die *Collision Induced Dissociation* (CID) – Methode durchgeführt. Von den aufgenommenen Profilspektren wurden durch das zum Massenspektrometer zugehörige Bearbeitungsprogramm Summenformeln bestimmt und anschließend Linienspektren zur Datenreduktion erstellt. Die m/z - Intensität Wertepaare wurden in TXT-Dateien gemeinsam mit Auflösung und Signal-zu-Rauschverhältnis abgespeichert. Die Informationen über die Summenformeln sind in CSV-Dateien abgelegt. Es liegen jeweils eine TXT – und eine CSV – Datei für die beiden Messungen vor. Zusätzlich wurde von der Fragmentierung ein MZXML-File erstellt, welches die m/z - und Intensitäts-Werte der einzelnen Scans in 0,5 eV Schritten beinhaltet.

Im Folgenden werden die Begriffe nicht fragmentiertes, fragmentiertes und Scan-Spektrum verwendet.

4.1.2 Analyt

Als Analyt diene der Nahrungsmittelzusatzstoff Saccharoseacetatisobutyrat²³ der zum Beispiel in der Getränkeherstellung als Stabilisator genutzt wird. **[41]** Dabei handelt es sich um ein Stoffgemisch, dessen Komponenten aus Isobuttersäure, Essigsäure und Saccharose synthetisiert werden. In Abbildung 4-1 ist schematisch die Herstellung von SAIB und dessen Struktur dargestellt. Im Gemisch können weiterhin alle möglichen Isomere zu der in Abbildung 4-1 dargestellte Struktur auftreten. Zusätzlich können durch den Herstellungsprozess bedingt Verbindungen wie zum Beispiel Oktaisobutyrat oder unterschiedliche Anzahl gebundener Isobutter- und Essigsäurereste²⁴ entstehen. **[42]**

²³ Abgekürzt: SAIB, E-Nummer E444

²⁴ Z.B. 5 Isobuttersäure- und 3 Essigsäure-Moleküle reagieren mit Saccharose

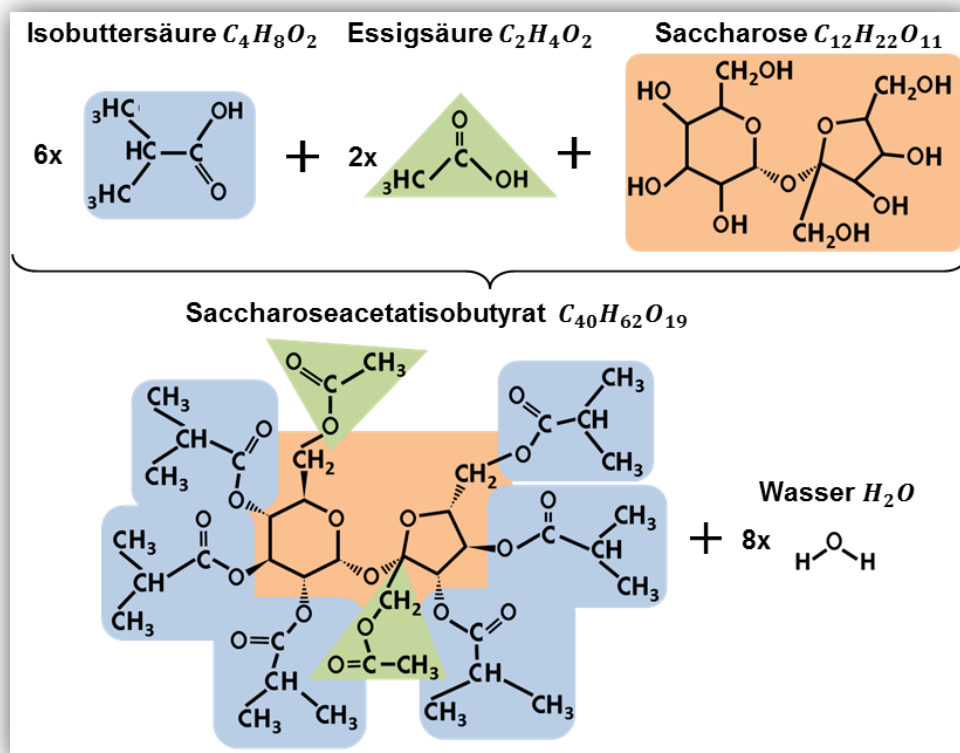


Abbildung 4-1: Herstellung von SAIB

Als Ausgangsstoffe dienen Isobuttersäure, Essigsäure und Saccharose. Bei der Synthese binden an Saccharose jeweils zwei Essigsäure- und sechs Isobuttersäurereste und es entsteht SAIB.

4.2 Datenformate

4.2.1 Aufbau der TXT-Datei

Die Textdateien speichern Peak-relevante Daten eines Massenspektrums und enthalten vier Spalten die jeweils durch einen Tabulator getrennt sind. Die erste Spalte enthält die m/z – Werte, die zweite die Intensität, die dritte die Auflösung und die vierte Spalte das Signal-zu-Rauschverhältnis eines Peaks wieder. Jede Zeile repräsentiert einen Peak im Spektrum. In Tabelle 4-1 ist der Aufbau schematisch dargestellt.

Tabelle 4-1: Aufbau der TXT-Dateien

Die Spalten sind von links nach rechts: m/z -Werte, Intensität, Auflösung und Signal-zu-Rausch Verhältnis.

Jede Zeile repräsentiert einen Peak im Spektrum

m/z	I	Res.	S/N
123.489357	2109083	512891	111.9
123.489494	1734992	1332615	88.9
123.489622	1653568	720127	83.9

Die Auflösung ist nach Formel 2.19 nach der FWHM-Definition angegeben. Da die Daten als Linienspektren gegeben sind, repräsentiert jedes Wertepaar einen Peak. Das nicht-fragmentierte Spektrum besitzt 391 Datenpunkte und das fragmentierte 1763 Datenpunkte.

4.2.2 Aufbau der CSV-Datei

Die CSV-Dateien bestehen aus mindestens acht Spalten. Sie speichern Summenformeldaten die zu bestimmten Peaks eines Spektrums zugeordnet werden konnten. Je mehr Elemente in den Summenformeln vertreten sind desto mehr Spalten enthält die Tabelle. Die zweite und dritte Spalte enthalten die gemessene Intensität und den m/z-Wert eines. Diese zwei Werte sind auch in der TXT-Datei wiederzufinden. Die vierte Spalte enthält den Wert für die exakte monoisotopische Masse des Moleküls der entsprechenden Summenformel. Die Massendifferenz zwischen berechnetem und gemessenem m/z-Wert ist in Spalte fünf in *Millidalton* (mDa) angegeben. Die Massengenauigkeit ist in Spalte sechs gespeichert und wird in *parts per million* (ppm) angegeben. Die siebente Spalte enthält die Summenformel des Moleküls. Alle folgenden Spalten sind abhängig von der Zusammensetzung der Moleküle. Jede Spalte repräsentiert ein Element, welche die jeweilige Anzahl jedes Elements in der jeweiligen Summenformel angibt. Kommt ein Element nicht in einer Summenformel vor so wird eine „0“ eingetragen. Die Spalten sind durch Tabulatoren voneinander getrennt. In Tabelle 4-2 ist der Aufbau schematisch dargestellt.

Tabelle 4-2: Aufbau der CSV-Datei

Die ersten sieben Spalten geben von links nach rechts das mSigma, gemessene Intensität und m/z-Verhältnis, berechnetes m/z-Verhältnis, Massendifferenz, Massengenauigkeit und Summenformel wieder.

Die folgenden Spalten enthalten die Anzahl der Elemente in den jeweiligen Summenformeln.

mSigma	Observed Intens	Observed m/z	calc. m/z	err mDa	err ppm	sum formula	C	H	O
12.60	70729776	197,080875	197,0808	-0,039	-0,2	C10 H13 O4	10	13	4
6.33	287603232	239,091442	239,0914	-0,042	-0,177	C12 H15 O5	12	15	5
8.80	145991718	267,122723	267,1227	-0,022	-0,084	C14 H19 O5	14	19	5

Die exakte Masse wird mit Hilfe von Formel 3.3 und die Massengenauigkeit durch Formel 2.20 berechnet.

4.2.3 Aufbau der mzXML-Datei

Das mzXML Format beinhaltet die Daten von Kollisions-Experimenten und basiert auf XML²⁵. Die für diese Arbeit wichtigen Informationen sind die m/z -Werte und Intensitäten aller vorhandenen Scans. Diese sind in dem Element `<peaks>` innerhalb der `<scan>` Elemente abgespeichert. In Abbildung 4-2 ist eine stark gekürzte Form der vorliegenden mzXML Datei dargestellt. Die Attribute `precision`, `byteOrder` und `pairOrder` der `<peaks>` Elemente geben die Art, nach der die m/z -Werte und Intensitäten abgespeichert sind, an. Es liegen insgesamt 41 Scans vor.

```
<mzXML xsi:schemaLocation="...">
  <msRun scanCount="41">
    ...
    <scan num="1" ...>
      <peaks precision="32" byteOrder="network" pairOrder="m/z-int">
        QwEODUtdQVIDAQ4gTNvH6EMBY79LiP3UQwpmA0s4kWhDCmYOTCYN
        8EMKZiVNyeBiQwq7xEyQPABDE74DS4OSYEMTvhBMakV3Qx0+K04LQp
        pDFBOvS1...
      </peaks>
    </scan>
    <scan num="2" ...>
      ...
    </scan>
    ...
  </msRun>
  ...
</mzXML>
```

Abbildung 4-2: MZXML Format

Abgebildet ist eine stark vereinfachte Form des mzXML Formats. Für die Arbeit wurden die m/z -Werte und die Intensitäten jedes Scans extrahiert.

Eine ausführliche Dokumentation des .mzXML Formats ist unter <http://www.peptideatlas.org/tmp/mzML1.1.0.html> verfügbar.

4.3 Software

Die Entwicklung der Oberfläche sowie Algorithmen zur Simulation der Spektren und Berechnung der Isotopenmuster wurde die Programmiersprache JAVA v. 1.7 verwendet. Als Entwickleroberfläche diente Eclipse Juno v. 4.2.2²⁶. Für die grafische

²⁵ eXtensible Markup Language

²⁶ Verfügbar unter: <http://www.eclipse.org/downloads/packages/eclipse-classic-422/junosr2>

Visualisierung der Spektren wurde die JFreeChart Bibliothek benutzt²⁷. Des Weiteren wurde das Serialisierungs-Tool XStream v. 1.4.7²⁸ verwendet.

²⁷ Verfügbar unter: <http://www.jfree.org/jfreechart/download.html>

²⁸ Verfügbar unter: <http://xstream.codehaus.org/>

5 SEAT

Als wesentlicher Teil dieser Arbeit wurde das Programm SEAT – *Spectrum Evaluation and Analyzation Tool* entwickelt (siehe beigelegte CD und Handbuch). Es dient der Analyse experimentell bestimmter Spektren und verwendet die hier vorgestellten Algorithmen. Die vorliegenden Daten sind im vorhergehenden Kapitel erläutert worden. In diesem Kapitel werden das Programm und dessen Algorithmen erläutert sowie anhand des Datensatzes erzeugte Ergebnisse. Zuvor wird jedoch kurz auf die wichtigsten Basisfunktionen von SEAT eingegangen. Eine Dokumentation über die Funktionen und Bedienung des Programmes ist im Handbuch zu finden.

5.1 Basisfunktionen von SEAT

5.1.1 Import von TXT, CSV und MZXML Dateien

Das Programm unterstützt die im vorherigen Kapitel vorgestellten Formate *.txt*, *.csv* und *.mzXML*. Die Daten müssen als Linienspektren vorliegen, d.h. jede Datenzeile bzw. *m/z*-Intensitäts - Wertepaar repräsentiert einen Peak im Spektrum. Die *.txt* und *.csv* Daten werden vollständig in das Programm übernommen. Von den *.mzXML* Dateien werden die *m/z*-Werte und die Intensitäten eingelesen.

Die Datenstruktur des Programmes ist in Workspaces organisiert. In Abbildung 5-1 ist das Datenmodell schematisch dargestellt. Jede Workspace beinhaltet ein oder mehrere Projekte, in denen importierte und simulierte Spektren gespeichert werden. Der Import von Dateien wird im Folgenden erläutert. *Simulierte Spektren* werden durch die in Kapitel 5.2 vorgestellten Algorithmen mit Hilfe von Summenformeldaten berechnet. Sie werden unter dem Punkt „*Simuliert*“ in einem Projekt abgespeichert.

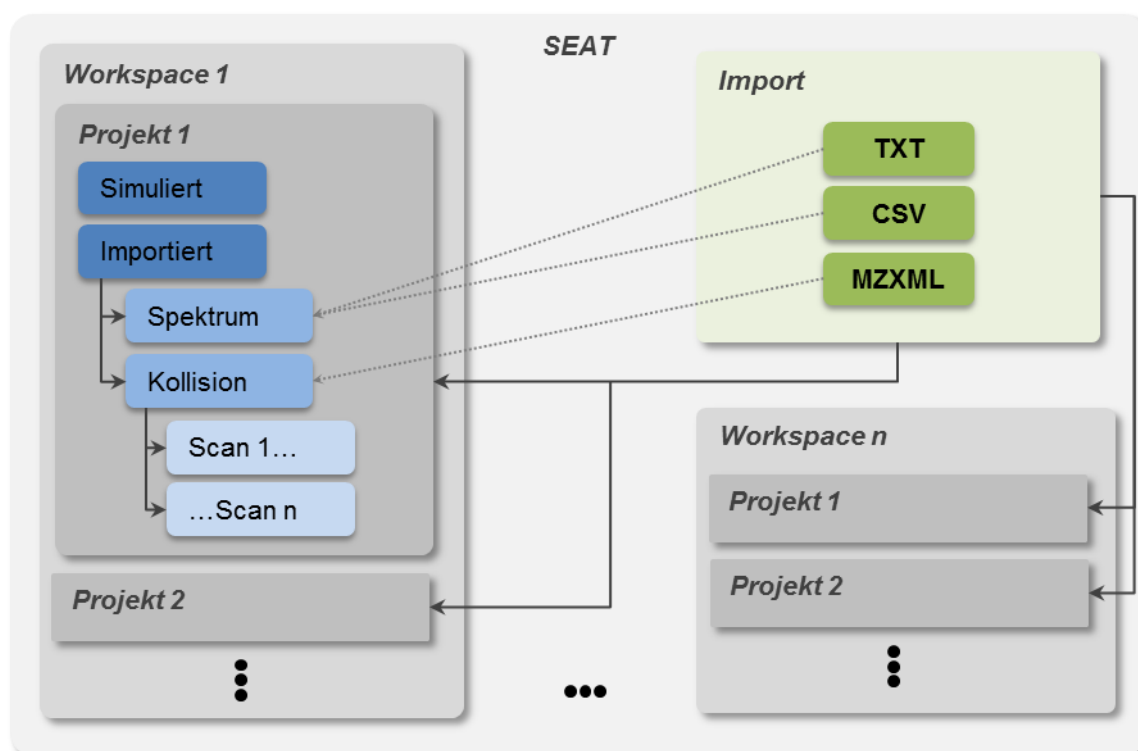


Abbildung 5-1: Datenmodell

SEAT importiert Linienspektren der Formate TXT, CSV und MZXML. Die in Workspaces organisierten Projekte enthalten die simulierten und importierten Spektren.

Jedes Projekt kann jeweils eine TXT, eine CSV und eine MZXML Datei importieren. In Abhängigkeit der gewählten Datenformate für den Import werden zwei verschiedene Datenmodelle für die Repräsentation der Daten unterschieden:

Das *Spektrum* repräsentiert die Daten von CSV und TXT Dateien. Programmintern wird jede einzelne Spalte der Datei als Liste oder Array repräsentiert. In Abbildung 5-2 ist dies schematisch dargestellt. Werden beide Formate importiert, so werden die Daten der CSV und TXT Datei, also Peak-Daten und zugehörige Summenformelinformationen automatisch miteinander verbunden. Dabei werden die m/z – Werte der TXT Datei mit den m/z_{meas} – Werten der CSV Datei zeilenweise miteinander verglichen. Stimmen die zwei Werte überein, so werden die entsprechenden Zeilen mit Hilfe eines Index miteinander verknüpft. In Abbildung 5-2 sind diese Übereinstimmungen schematisch durch rot und blau markierte Zeilen gekennzeichnet.

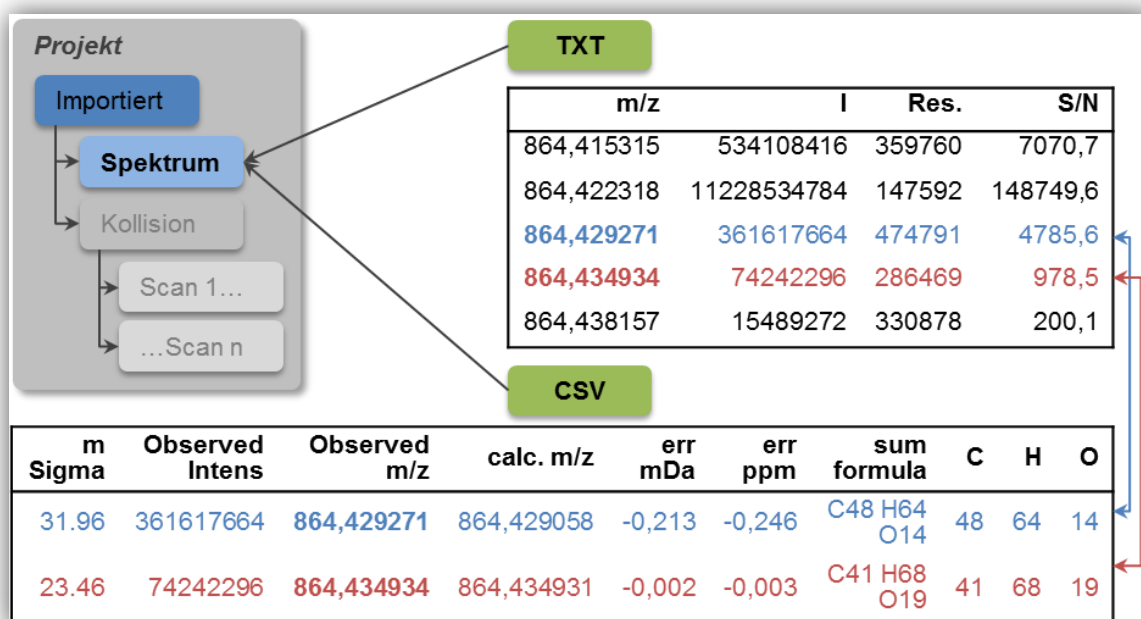


Abbildung 5-2: Import von TXT und CSV Datei

Werden TXT und/oder CSV Dateien importiert, werden diese in einem Projekt als *Spektrum* repräsentiert. Liegt sowohl eine TXT als auch eine CSV Datei vor, werden die Daten durch einen zeilenweisen Vergleich der m/z - Werte der TXT und m/z_{meas} - Werte der CSV miteinander verknüpft. Ist ein m/z -Wert gleich einem m/z_{meas} - Wert so korrespondieren die entsprechenden Daten. Dies ist in der Abbildung durch die rote bzw. blaue Markierung der Zeilen verdeutlicht.

Unter *Kollision* werden importierte MZXML Daten abgespeichert. Da eine MZXML Datei mehrere Scans enthält, wird jeder Scan einzeln in das Projekt geladen und angezeigt. Es ist ebenfalls möglich die *Scans* mit Summenformeldata zu verknüpfen. Da jedoch die m/z Werte der einzelnen Scans variieren, müssen diese vorher aligniert werden um später eine Entscheidung über die Zugehörigkeit der Summenformelinformationen treffen zu können. Dies ist in Abbildung 5-3 angedeutet. Beginnend beim kleinsten m/z - Wert m/z_{min} aus allen Scans wird mit Hilfe eines definierten ppm-Schwellwerts ppm_T das Intervall I mit

$$I = [m/z_{min}, m/z_{min} + \Delta(m/z)]$$

$$\text{mit } \Delta(m/z) = \frac{ppm_T \cdot m/z_{min}}{10^6} \text{ und } m/z_{min} = \min(m/z) \quad (5.1)$$

Berechnet. Alle m/z - Werte der Scans innerhalb dieses Intervalls werden somit aligniert. Anschließend wird der nächstgrößere m/z - Wert außerhalb dieses Intervalls gesucht, ein neues Intervall berechnet und erneut eine Zuordnung vorgenommen. Dies wird solange wiederholt bis alle Peaks aller Scans aligniert wurden. Das Ergebnis ist eine Matrix wie in Abbildung 5-3 (obere Tabelle) dargestellt in der bildlich gesehen die alignierten Peaks eines Intervalls in einer Zeile der Matrix gespeichert sind. Für

jede Zeile wird nun der Mittelwert $mean(\{m/z\}_I)$ berechnet und die ppm – Abweichung zu den m/z_{meas} – Werten der CSV Datei ermittelt. Die Daten des Wertepaars mit der geringsten ppm-Abweichung werden anschließend mit allen Scans verknüpft. Es kann vorkommen, dass einige Summenformelinformationen nicht zu allen Scans zugeordnet werden können.

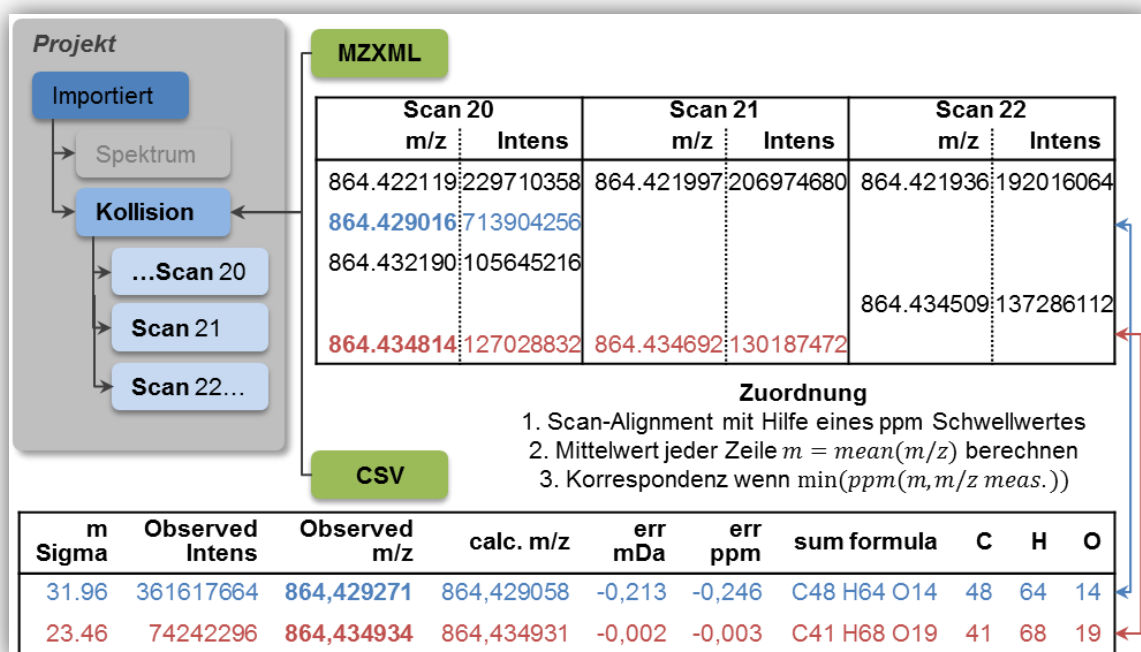


Abbildung 5-3: Import von MZXML und CSV

MZXML und CSV Daten werden ebenfalls miteinander verknüpft und unter *Kollision* in einem Projekt gespeichert. Jeder Scan erhält seinen eigenen Eintrag, da es sich um einzelne Spektren handelt. Die Summenformelzuordnung erfolgt nach Scan-Alignment mit Hilfe der geringsten ppm-Abweichung.

Ein Projekt kann sowohl ein *Spektrum* als auch eine *Kollision* enthalten. Im Folgenden weisen die kursiv geschriebenen Wörter „*Spektrum*“, „*Scan*“ und „*Kollision*“ auf die Datenstrukturen in einem Projekt hin.

5.1.2 Approximierte Auflösung für Spektren und Scans

Eine wichtige Eigenschaft von Peaks ist die Auflösung. In dieser Arbeit wird sie für verschiedene Probleme als Parameter benötigt: Für die Berechnung eines simulierten Spektrums mit der 2D-FT Methode, der Darstellung der Spektren oder Scans als approximiertes Profilspektrum oder dem Vergleich von Spektren.

Da jedoch nur TXT Dateien die Auflösungen der Peaks speichern, fehlen diese bei einem Import von CSV und MZXML Dateien. Aus diesem Grund bietet SEAT die

Möglichkeit, für ein *Spektrum* oder einen *Scan* eine approximierte Auflösung auf Basis von m/z - Werten zu berechnen. Dafür stehen drei Funktionen zur Verfügung:

1. Die lineare Funktion f_L mit

$$f_L(m/z) = \frac{R}{(m/z)_R} \cdot m/z \quad (5.2)$$

2. Die Exponentialfunktion 1 f_{E1} mit

$$f_{E1}(m/z) = \nu \cdot e^{\frac{m/z}{(m/z)_R} \cdot \ln\left(\frac{R}{\nu}\right)} \quad \text{mit } \nu = 10^{\lceil \log(R) \rceil} \quad (5.3)$$

3. Die Exponentialfunktion 2 f_{E2} mit

$$f_{E1}(m/z) = a \cdot e^{b \cdot (m/z)} + c \cdot e^{d \cdot (m/z)} \quad (5.4)$$

Die Funktionen f_L und f_{E1} benötigen als Parameter eine Auflösung R bei dem m/z -Wert $(m/z)_R$. Wird für $(m/z)_R = 0$ gewählt wird Funktion f_L und f_{E1} zu einer konstanten Funktion $f(m/z) = R$ umgewandelt. Die Funktion f_{E2} benötigt die Parameter $a - d$.

Für die Approximation der Auflösung eines FT-ICR-Massenspektrometers ist die Funktion f_{E2} durch die Kombination von zwei Exponentialfunktionen und der dadurch erhöhten Genauigkeit am besten geeignet. Die Parameter $a - d$ können entweder frei gewählt oder berechnet werden. Die Standardwerte, die das Programm SEAT für die Parameter $a - d$ bereitstellt, wurden mit Hilfe der *Curve – Fitting Toolbox* von *Matlab*²⁹ berechnet. Jedes andere Programm welches exponentielles Curve - Fitting unterstützt ist jedoch genauso gut geeignet. Als Daten wurden die m/z – Werte und die zugehörigen Auflösungen einer TXT - Datei eines experimentell bestimmten Massenspektrums genutzt. In Abbildung 5-4 ist das Ergebnis des Curve-Fits dargestellt.

²⁹ Es wurde die Version MATLAB 7.12.0.635 (R2011a) und die Curve Fitting Toolbox 3.1 verwendet. Das Curve Fitting erfolgte mit den Einstellungen: Type of Fit = Exponential $a \cdot \exp(b \cdot x) + c \cdot \exp(d \cdot x)$, Fit Options: Robust = LAR, Algorithm = Trust-Region.

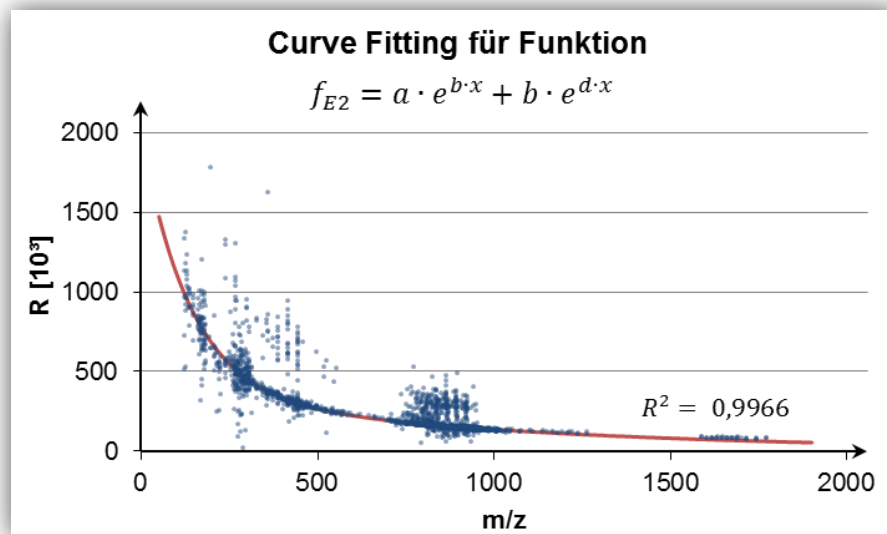


Abbildung 5-4: Ergebnis des Curve-Fitting eines experimentellen Massenspektrums

Es wurden die Auflösungen in Abhängigkeit der m/z -Werte eines experimentellen Massenspektrums aufgetragen (blaue Punkte). Die Funktion f_{E2} (rote Linie), welcher mit der Curve-Fitting Toolbox von MatLab erstellt wurde, repräsentiert den Verlauf der Auflösung sehr gut. Die ermittelten Werte der Parameter sind $a = 1612000$, $b = -0,00732$, $c = 373400$ und $d = -0,001015$.

Der Vorteil der f_{E2} Funktion liegt darin, dass die berechneten Auflösungen gut an reale Werte angenähert werden können. Dies ist vor allem bei dem Vergleich von Spektren basierend auf m/z -Wert und Auflösung hilfreich.

In Abbildung 5-5 sind alle drei Auflösungs-Funktionen mit den in SEAT implementierten Standardwerten ($(m/z)_R = 400$, $R = 330.000$ sowie $a = 1612000$, $b = -0,00732$, $c = 373400$ und $d = -0,001015$ dargestellt.

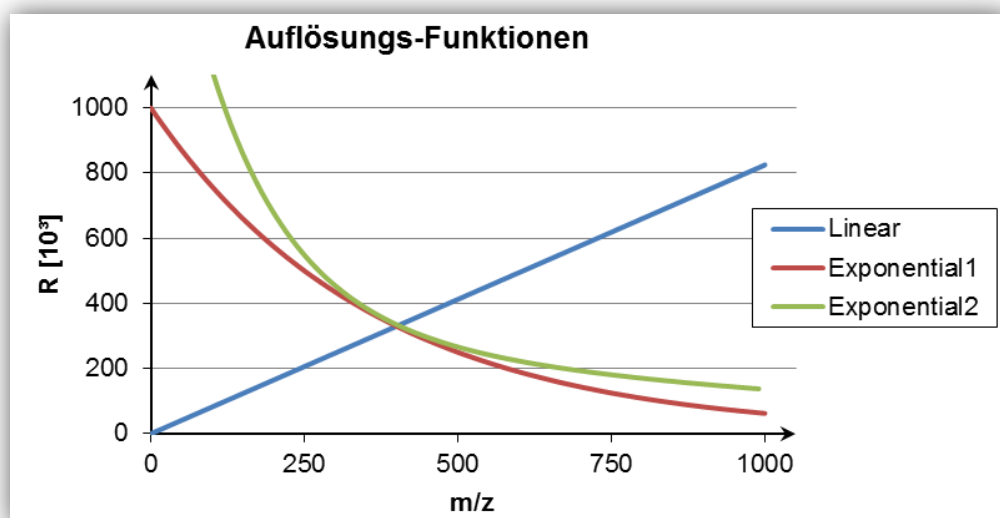


Abbildung 5-5: Funktionen für die Approximation der Auflösung von Spektren

Es sind die drei Funktionen Linear (blau), Exponential1 (rot) und Exponential2 (grün) mit den in SEAT implementierten Standardwerten abgebildet. Es ist jeweils die Auflösung in Abhängigkeit von den m/z -Werten dargestellt.

5.1.3 Grafische Darstellung

SEAT stellt zwei verschiedene Darstellungsmöglichkeiten der Spektren zur Verfügung. Das Linienspektrum und das Profilspektrum. Im Profilmodus kann sowohl die Gauß-Funktion als auch die Lorentz-Funktion als Peak-Form gewählt werden. Für die Profilspektren werden vorrangig die gemessenen Auflösungen verwendet. Sind diese nicht vorhanden wird die eingestellte approximierte Auflösung berechnet. Für das Profilspektrum werden aggregierte Peaks berechnet. D.h. befinden sich mehrere Peaks in einem bestimmten m/z – Intervall I_m , welches dynamisch anhand der Auflösungen der Peaks berechnet wird, werden alle Intensitäten der Peak-Kurven summiert. Pro Peak werden 32 Datenpunkte verwendet, so dass jeder aggregierte Peak $32 \cdot n$ Datenpunkte besitzt. Wobei n die Anzahl der Peaks innerhalb des Intervalls I_m darstellt. Es können sowohl die absoluten als auch die relativen Intensitäten³⁰ genutzt werden. Des Weiteren gibt es die Möglichkeit, je nach Bedarf die Summenformeln im Spektrum anzeigen oder ausblenden zu lassen. In Abbildung 5-6 ist ein Ausschnitt aus dem nicht fragmentierten Spektrum von SAIB als Profilspektrum mit gaußscher Peak-Form dargestellt.

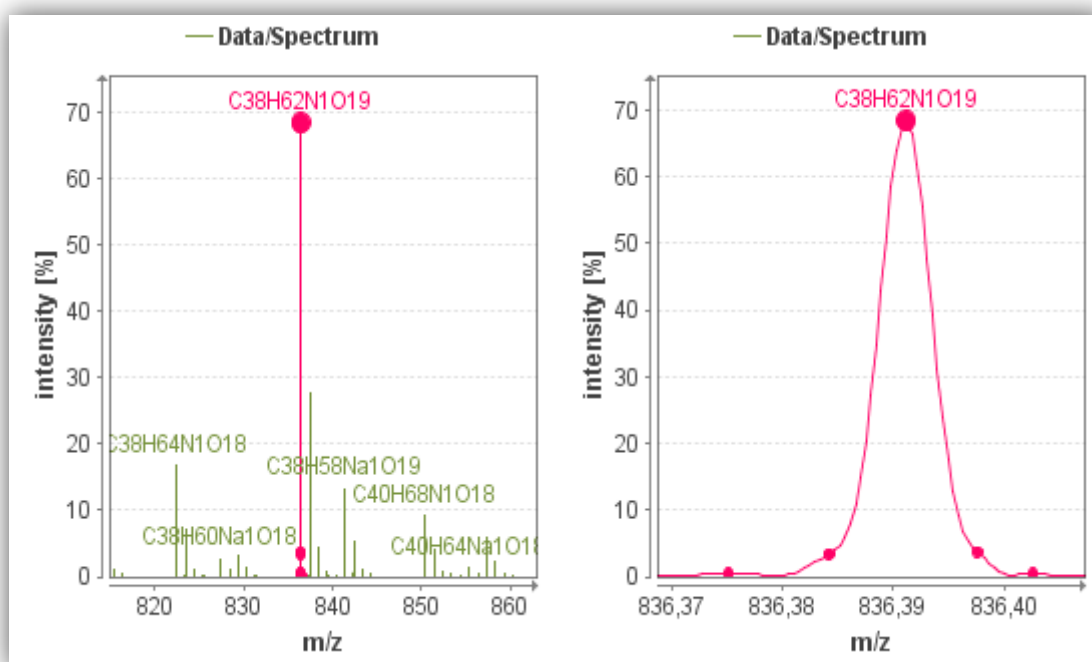


Abbildung 5-6: Darstellung des nicht fragmentierten Spektrums als Profilspektrum

Links: Dargestellt ist ein Ausschnitt aus dem nicht fragmentierten Spektrums von SAIB als Profilspektrum mit gaußscher Peak-Form. Der markierte Peak ist rechts vergrößert. **Rechts:** Ein aggregierter Peak aus fünf Peaks. Die einzelnen Peaks sind durch Kreise hervorgehoben. Ist zu einem Peak eine Summenformel assoziiert ist der Kreis etwas größer.

³⁰ Normierung auf 100 %

5.2 Berechnung simulierter Spektren

Als Voraussetzung für die Simulation von Spektren muss mindestens eine Summenformel in einem *Spektrum* oder einem *Scan* einer *Kollision* gegeben sein. Weitere Voraussetzungen werden bei der jeweiligen Methode erläutert. Für jede Summenformel wird das Isotopenmuster berechnet. Dafür wurden zwei Algorithmen implementiert: die multinomiale nach Yergey [35] und eine auf Fourier-Transformation basierte Methode. Für letzteres wurde ein Programmpaket namens Deuterium verwendet [40]. Liegen alle Isotopenmuster vor, so werden diese summiert, so dass ein komplettes simuliertes Spektrum entsteht.

5.2.1 Multinomiale Methode

Die Herleitung dieses Algorithmus wurde im Kapitel 3.4.2 erläutert. In diesem Abschnitt werden die Algorithmischen Details erläutert.

Der erste Schritt bei der Berechnung der Isotopenmuster mit der multinomialen Methode besteht darin alle möglichen Isotopenkombinationen einer Summenformel zu finden. Die grundlegenden Schritte bestehen aus:

1. Berechnen aller möglichen Isotopenverteilungen jedes Elements der Summenformel
2. Berechnen der Häufigkeiten der Isotopenverteilungen
3. Multiplikation der Isotopenverteilungen der Elemente
4. Berechnung der Masse jeder Isotopenverteilung

Die Berechnung der Isotopenverteilungen kann elementweise durchgeführt werden. Für jedes Element in der Summenformel S sind i Isotope und n Atome als Variablen vorhanden. Die Bestimmung aller Isotopenverteilungen eines Elements ist äquivalent mit dem mathematischen Problem der Bestimmung aller Partitionen einer ganzen positiven Zahl und allen Permutationen aus diesen Partitionen. Eine Partition ist definiert als die geordnete Summe

$$n = \sum_{j=1}^i n_j \quad (5.5)$$

von positiven ganzen Zahlen $n_1 \dots n_j$. [43] Sie kann in aufsteigender oder absteigender Ordnung angegeben werden. Zum Beispiel sind die Partitionen für $n = 4$ 1111, 112, 22, 13 und 4. Die Analogie zur Berechnung von Isotopenverteilungen wird

hergestellt indem man annimmt, dass der Index j die jeweiligen Isotope und die Zahl n_j die Anzahl der Isotope eines Elements in einer Isotopenverteilung repräsentiert. Bei dieser Annahme muss man jedoch beachten, dass alle Partitionen mit mehr als i Summanden für die Berechnung der Isotopenverteilungen wegfallen. Zum Beispiel würde bei C_4 nur 40, 31 und 22 für weitere Berechnungen verwendet werden, da Kohlenstoff nur zwei Isotope besitzt. Für jedes Element a, b, \dots in der Summenformel S werden nun alle Partitionen $N = \{(n)_1 \dots (n)_p\}$ und deren Permutationen $P = \{p((n)_1) \dots p((n)_p)\}$ gesucht. In Tabelle 5-1 sind am Beispiel von O_3 die Partitionen in der zweiten Spalte (300, 210 und 111) und deren Permutationen in den darauffolgenden Spalten dargestellt.

Tabelle 5-1: Partitionen und deren Permutationen am Beispiel von O_3

In der Tabelle sind die Partitionen und deren Permutationen von Ozon dargestellt. Ozon besteht aus drei Sauerstoffatomen ($n = 3$). Sauerstoff besitzt drei Isotope ($k = 3$). Jeder Wert der Variable i repräsentiert ein Isotop von Sauerstoff. Die drei Partitionen sind in der zweiten Spalte und deren Permutationen in den darauffolgenden Spalten dargestellt. Die „|“ sind nur zur besseren Übersicht eingefügt.

	Partitionen			Permutationen		
i	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3	1 2 3
Isotope	$^{16}O ^{17}O ^{18}O$	$^{16}O ^{17}O ^{18}O$	$^{16}O ^{17}O ^{18}O$	$^{16}O ^{17}O ^{18}O$	$^{16}O ^{17}O ^{18}O$	$^{16}O ^{17}O ^{18}O$
$p((n)_1)$	3 0 0	0 3 0	0 0 3			
$p((n)_2)$	2 1 0	2 0 1	1 2 0	1 0 2	0 1 2	0 2 1
$p((n)_3)$	1 1 1					

Algorithmus 5-1 beschreibt die Berechnung der Partitionen und deren Permutationen. Als Grundlage dafür wurde der Algorithmus 4.1 aus [43] verwendet. Es werden die zwei Übergabeparameter i und n benötigt, die jeweils die Anzahl der Isotope des Elements bzw. die Anzahl der Atome des Elements als ganze positive Zahlen repräsentieren. Jede Partition und Permutation wird als Array der Länge i abgespeichert. Diese werden wiederum in einem Array *permutationen* zusammengefasst und als Ergebnis des Algorithmus zurückgegeben.

Der Algorithmus ist im Wesentlichen in drei Teile aufgegliedert. Im ersten Teil, welcher die ersten beiden *if*-Verzweigungen umfasst werden monoisotopische und diisotopische Elemente behandelt. Handelt es sich um ein monoisotopisches Element wird n zurückgegeben, da es hier nur diese eine mögliche Isotopenverteilung gibt. Bei diisotopischen Elementen werden mit Hilfe einer einfachen *for*-Schleife die zwei-

elementigen Arrays aufgefüllt. Aufgrund dieser Vorbehandlung mono- und diisotopischer Elemente wird der Algorithmus, vor allem bei größeren Werten von n beschleunigt.

Der zweite Teil umfasst die Berechnung der Partitionen für polyisotopische Elemente. Das Array a wird für die Generierung der Partitionen verwendet. Es besteht aus ganzen Zahlen. Bei der Initialisierung wird a mit dem Wert „0“ gefüllt. Die Variablen k und l dienen als Indices für das Array a und x und y repräsentieren die Werte die in das Array an den Positionen l oder k geschrieben werden. Um nur die Partitionen abzuspeichern die eine maximale Länge von i besitzen, wurde der ursprüngliche Algorithmus aus [43] abgewandelt und die beiden *if*-Verzweigungen eingefügt. Damit alle Partitionen die gleiche Länge i besitzen, werden sie, falls nötig, mit „0“ aufgefüllt und nach absteigender Ordnung sortiert. Dies erleichtert die folgenden Rechenschritte. Jede Partition wird anschließend an den Algorithmus zur Berechnung aller Permutationen, `berechnePermutationVonPartition`, weitergegeben. Dieser ist in Algorithmus 5-2 dargestellt. Als Ergebnis erhält man alle Permutationen zusammengefasst in dem Array b welches der Rückgabeveriable `partitionen` hinzugefügt wird.

Algorithmus 5-1: Berechnen der Partitionen

Der Parameter i entspricht der Anzahl der Isotope und n der Anzahl der Atome eines Elements. Für jede Partition werden zusätzlich die Permutationen berechnet um alle möglichen Isotopenverteilungen zu erhalten.

```

BerechnePartitionen(Integer i, n)
    Integer[][] partitionen;
    if( i = 1 )
        permutationen = {n}
        return partitionen
    if( i = 2 )
        for( h = n, h ≥ 0, h = h - 1 )
            Integer[] partition
            partition[0] = h
            partition[1] = n - 1
            partitionen.add(partition)
        return partitionen
    else
        Integer[] a = {0}
        k = 1
        y = n - 1
        while( k != 0 )
            k = k - 1
            x = a[k] + 1
            while( 2 · x ≤ y )
                a[k] = x
                y = y - x
                k = k + 1
            l = k + 1
            while( x ≤ y )
                a[k] = x
                a[l] = y
                if( l < i )
                    Integer[] partition = a[0 ... l]
                    partition[l + 1 ... i - 1] = 0
                    sortiere partition in absteigender Ordnung
                    b [][] = berechnePermutationenVonPartition(partition)
                    partitionen.add( b )
                x = x + 1
                y = y - 1
            y = y + x - 1
            a[k] = y + 1
            if( k < i )
                partition = a[0 ... k]
                partition[k + 1 ... i - 1] = 0
                sortiere partition in absteigender Ordnung
                b [][] = berechnePermutationenVonPartition(partition)
                partitionen.add( b )
        return partitionen

```

Die Anzahl aller Permutationen einer Partition der Länge i mit verschiedenen Elementen ergibt sich durch $i!$. Gibt es in einer Partition doppelte Zahlen ist die Anzahl mit $i! / p_1 + \dots + p_m$, $i = p_1 + \dots + p_m$ bestimmbar. In Algorithmus 5-2 werden die einzelnen Permutationen einer Partition Zahl für Zahl berechnet. Gegeben ist die Partition $n_1 \dots n_j$ repräsentiert durch das Array *partition* von Algorithmus 5-2. Für die Initialisierung werden die ersten beiden Zahlen $n_1 n_2$ und $n_2 n_1$ als Teilarray in dem Array *permutationen* gespeichert. Die Permutationen werden schrittweise erzeugt. Ausgehend von den Teilpartitionen $n_1 n_2$ und $n_2 n_1$ wird in jedem Iterationsschritt eine Zahl z nach dem Schema von Formel 5.6 an jede Stelle l in der Teilpartition eingefügt.

$$[n_1 n_2] \xrightarrow{n_3} [n_3 n_1 n_2] [n_1 n_3 n_2] [n_1 n_2 n_3] \xrightarrow{n_4} \dots \xrightarrow{n_j} \quad (5.6)$$

Dies wird für alle Teilpartitionen die in dem Array *temp* als Zwischenergebnis gespeichert sind wiederholt. Am Ende eines Iterationsschritts sind alle Arrays in *temp* um eine Zahl länger als davor. Alle Permutationen sind berechnet wenn die letzte Zahl n_j der der gegebenen Partition *partition* nach Formel 5.6 verarbeitet wurde.

Algorithmus 5-2: Berechne Permutationen von einer Partition

Der Algorithmus berechnet alle Permutationen der Partition *partition*. Ausgehend von den ersten beiden Teil-Arrays *part1* und *part2* werden die Permutationen in jedem Iterationsschritt Zahl für Zahl aufgebaut.

```
BerechnePermutationenVonPartitionen(Integer[] partition)
Integer[][] permutationen;
part1 = partition[0 ... 1]
permutationen.add( part1 )
part2 = partition[1 ... 0]
permutationen.add( part2 )
for( i = 2, i < partition.size, i = i + 1 )
    z = partition.get( i )
    Integer[][] temp
    for( k = 0, k < permutationen.size, k = k + 1 )
        p[] = permutation[k]
        for( l = 0, l < i + 1, l = l + 1 )
            q = p
            q[l] = z
            if( !( permutationen.contains( q ) || temp.contains( q ) ) )
                temp.add( q )
        permutationen = temp
    temp.clear
return permutation
```

Liegen alle Permutationen aller Partitionen, d.h. alle möglichen Isotopenverteilungen vor kann die Häufigkeit dieser berechnet werden. Als Voraussetzung dafür benötigt

man zunächst die Isotopenzusammensetzung der Elemente. Diese wurden aus dem IUPAC Technical Report 2009 „*Isotopic Composition of the Elements*“ [26] entnommen. Zusätzlich zu weiteren Informationen sind diese Daten in Tabelle A-1 zu finden. Für die Berechnung wurde die relative Isotopenzusammensetzung verwendet. Die Berechnung der Häufigkeiten ist mit Formel 3.10 und Formel 3.11 möglich. Durch die Berechnung der Fakultäten entstehen sehr schnell sehr große Zahlen im Zähler bzw. Nenner des Bruchs. Wird bei der Programmierung einer Datentyp doppelter Präzision (double) gewählt beträgt der Wertebereich $\pm 1,79769313486231570 \cdot 10^{308}$. Dies bedeutet das eine maximale Fakultät von $170! = 7,257415615307994 \cdot 10^{306}$ berechnet werden kann und damit alle Moleküle maximale 170 Atome pro Element besitzen dürfen unter Verwendung von Formel 3.10.

Von Yergey wurde Formel 3.11 für eine Reduktion an Multiplikationen durch Kürzen innerhalb des Quotienten vorgeschlagen. Für die Berechnung der Häufigkeiten der Isotopenverteilungen nach der multinomialen Methode wurde in dieser Arbeit ein anderer Ansatz gewählt:

Man kann jeden einzelnen Summanden aus Formel 3.10 mit der die Häufigkeit einer bestimmte Isotopenverteilung $n_1 \dots n_k$ berechnet werden kann, auch definieren als

$$H_n = e^{\ln\left(\frac{n!}{n_1! \dots n_k!} (a_1)^{n_1} (a_2)^{n_2} \dots (a_k)^{n_k}\right)} \quad (5.7)$$

Durch die Anwendung der Logarithmengesetze können so die Multiplikationen in Summationen umgewandelt werden und es wird mit kleineren Zahlen gerechnet. Es entsteht folgender Ausdruck für den logarithmischen Term

$$\ln(n!) - (\ln(n_1!) + \dots + \ln(n_k!)) + \ln((a_1)^{n_1}) + \dots + \ln((a_k)^{n_k}) \quad (5.8)$$

wobei für jede logarithmierte Fakultät gilt

$$\ln(n!) = (\ln(n) + \ln(n-1) + \dots + \ln(2) + \ln(1)). \quad (5.9)$$

Durch die Verwendung des natürlichen Logarithmus wird nur noch Formel 3.10 zur Berechnung der Häufigkeit benutzt. In Algorithmus 5-3 ist die Berechnung der Häufigkeit der zuvor berechneten Permutationen dargestellt. Als Parameter werden die beiden Arrays *permutationen* und *isotopenhfkt* benötigt, welche jeweils die Isotopenverteilung und die relativen Häufigkeiten der Isotope enthalten. Um die logarithmierten Fakultäten nicht immer neu berechnen zu müssen ist es sinnvoll zu Beginn der Isotopenmusterberechnung ein Array *logFakultaet* zu erstellen. Darin werden alle logarithmierten Fakultäten für $n = 1 \dots n_{max}$ gespeichert, wobei n_{max} der

Anzahl an Atomen des Elements in einer Summenformel entspricht das den höchsten Wert aufweist. In Anlehnung zu dem beschriebenen Algorithmus von Yergey in Kapitel 3.4.2 bestimmt ein Schwellwert darüber ob eine Permutation bzw. Isotopenverteilung und dessen berechnete Häufigkeit für weitere Berechnungen abgespeichert wird. Er errechnet sich aus der häufigsten Isotopenverteilung³¹ multipliziert mit einem Faktor *threshFaktor*. Yergey wählte in seiner Arbeit einen Wert von 0,01 [35]. Je kleiner der Wert desto mehr Peaks werden berechnet und desto mehr werden im simulierten Spektrum enthalten sein.

Algorithmus 5-3: Berechnung der Häufigkeiten der Isotopenverteilungen

Das Array *permutation* enthält alle berechneten Permutationen, *isotopenhfkt* die Häufigkeit der Isotope, und *logFakultaet* die vorher berechneten logarithmischen Fakultäten. Mit Hilfe dieser Transformation wird die Entstehung zu großer Zahlen verhindert. Wichtig ist die Rücktransformation mit Hilfe der Exponentialfunktion (*exp*) am Schluss der Berechnung. Die Bezeichnung *log* steht für den natürlichen Logarithmus. Es wird ein Array welches die Häufigkeiten enthält zurückgegeben.

```
BerechneHaeufigkeiten(Integer[][] permutationen, Integer[] isotopenhfkt, Integer atomAnzahl,
                        Double threshFaktor)

Double[] hfkten
threshhold = threshFaktor * max(isotopenhfkt)^ atomAnzahl
for( i = 0, i < partition.size; i = i + 1 )
    hfkt = 0;
    zaehler = 0
    nenner = 0
    potenzen = 0
    for( i = 0, i < permutation.size, i = i + 1 )
        zaehler = zaehler + permutation[i]
        nenner = nenner + logFakultaet[permutation[i]]
        potenzen = potenzen + log(isotopenhfkt[i]^permutation[i])
    hfkt = logFakultaet[zaehler] - nenner + potenzen
    hfkt = exp( hfkt )
    if( hfkt > threshhold )
        hfkten.add( hfkt )
return hfkten
```

Als Ergebnis von Algorithmus 5-3 erhält man die Häufigkeiten $H_A = \{H_{p_1}, H_{p_2}, \dots\}$ der Permutationen $P_A = \{p_1, p_2, \dots\}$ des Elements A . Im Anschluss an die Berechnung aller Häufigkeiten $H_S = \{H_A, H_B, H_C, H_D, \dots\}$ der Elemente A, B, C, D, \dots einer Summenformel $S = A_n B_m C_o D_p \dots$ folgt die Kombination dieser um alle Isotopenverteilungen und Häufigkeiten der Summenformel S zu erhalten. Dazu werden alle elementaren Häufigkeiten aller Elemente miteinander multipliziert. Es ist ebenfalls möglich mit Hilfe

³¹ Die häufigste Isotopenverteilung berechnet sich immer aus der relativen Häufigkeit des häufigsten Isotops potenziert mit der Anzahl der Atome des Elements.

eines Schwellwerts eine weitere Filterung der Summenformeln mit zu geringer Häufigkeit, genau wie in Algorithmus 5-3 vorzunehmen. Dazu kann der gleiche Wert *threshhold* verwendet werden. Addiert man alle Häufigkeiten einer Summenformel sollte die Summe eins ergeben. Da durch Verwendung von Schwellwerten einige Isotopenverteilungen nicht im Endergebnis enthalten ist die Summe meist geringfügig kleiner als eins. Um das simulierte Spektrum mit einem experimentellem vergleichen zu können, werden die Häufigkeiten auf eins normiert und mit den gemessenen Intensitäten multipliziert.

Der letzte Schritt in der Bestimmung des Isotopenmusters eines Moleküls besteht in der Berechnung der m/z -Verhältnisse der ermittelten Isotopenverteilungen. Die exakten Massen der Isotope wurden aus dem IUPAC Technical Report 2009 „*Atomic Weights of the Elements*“ [30] entnommen und sind in Tabelle A-1 aufgeführt. Da bei der Massenspektrometrie immer Ionen detektiert werden, wird das m/z -Verhältnis nach Formel 3.5 berechnet. Die Ladung eines Moleküls kann von dem experimentellen Spektrum bestimmt werden. Je höher die Ladung desto geringer der Abstand der Isotopenpeaks eines Moleküls. Die Werte für $c_1 \dots c_k$ in Formel 3.4 entsprechen den berechneten Permutationen der Elemente. Es ist daher wichtig alle Informationen zu Permutationen, Häufigkeiten und Elementen immer in Relation zueinander abzuspeichern. Die Werte der Isotopenmassen sind in Tabelle A-1 aufgeführt.

Als Endergebnis der Berechnung der Isotopenmuster nach der multinomialen Methode liegen folgende Daten für eine Summenformel S vor: Isotopenverteilungen, Häufigkeit/Intensität und m/z -Verhältnis jeder Isotopenverteilung, Summenformel mit Isotopeninformation. Das Schema in Abbildung 5-7 fasst den gesamten Algorithmus kurz zusammen. Das experimentelle Spektrum kann beliebig viele Summenformeln enthalten.

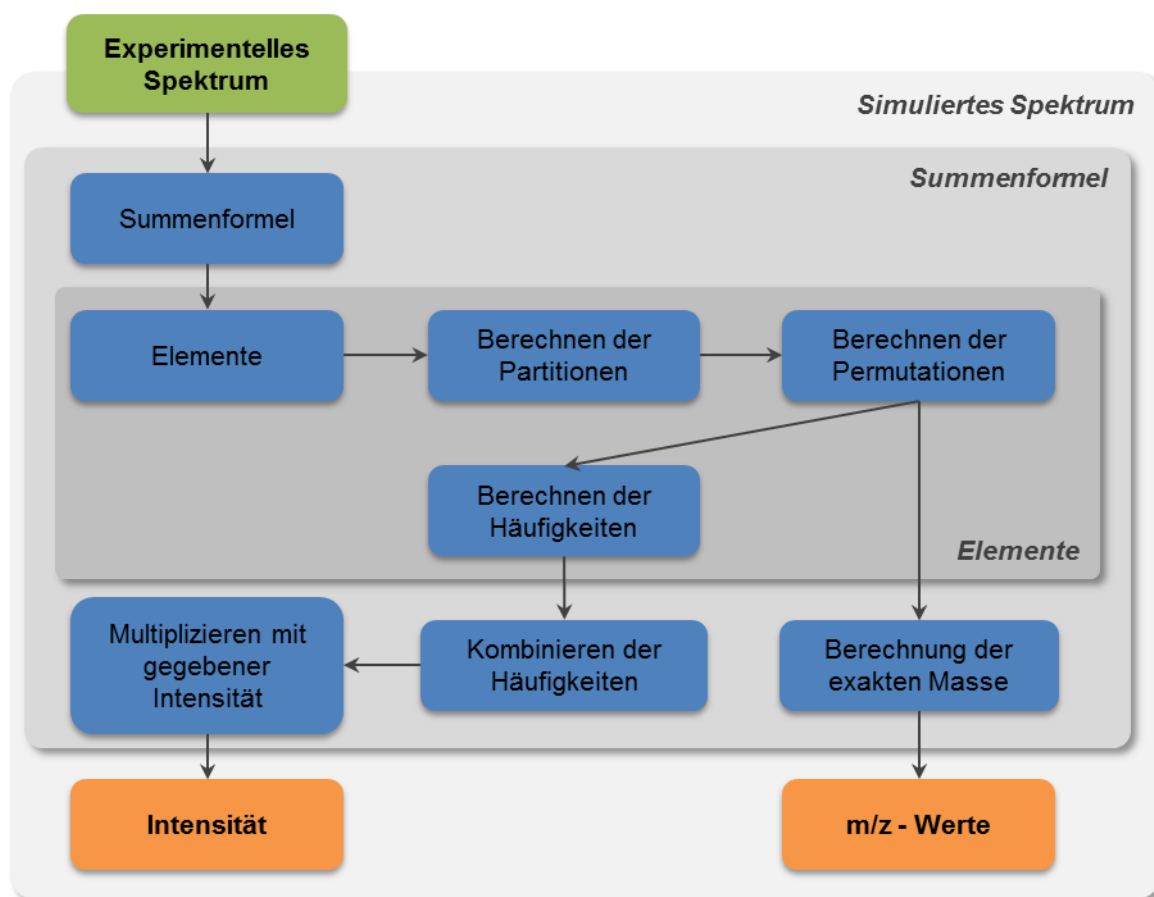


Abbildung 5-7: Schema der Berechnungsschritte der multinomialen Methode

Es ist der Verlauf der Berechnung eines simulierten Spektrums dargestellt. Ausgehend von den experimentellen Daten, die die Intensität sowie die Summenformeln für die Berechnung bereitstellen, wird jede Summenformel einzeln berechnet. Für jede Summenformel werden zuerst alle Häufigkeiten elementweise berechnet, diese dann kombiniert und mit der gemessenen Intensität multipliziert. Anhand der Isotopenverteilungen kann im Anschluss das m/z -Verhältnis ermittelt werden.

5.2.2 2D Fourier Transformation - *Deuterium*

Wie schon in der Einleitung dieses Kapitels erwähnt, wurde für die Berechnung der Isotopenmuster mit Hilfe der 2D-FT das Programm *Deuterium* [44] verwendet. Die Grundlagen zu diesem Algorithmus und dessen Herleitung wurde in Kapitel 3.4.3 und 3.4.4 beschrieben. Es berechnet Profilspektren eines Moleküls mit einer Auflösung bis zu 10^8 . Die Peak-Form entspricht einer Gauss-Kurve nach Formel 2.22. Dadurch, dass dieser Algorithmus kein Pruning wie bei der multinomialen Methode durchführt, um die Rechenschritte zu reduzieren, sind die Häufigkeiten, vor allem bei aggregierten Peaks

genauer. Weiterhin ist es möglich die Isotopenverteilungen auch von sehr großen Molekülen³² zu berechnen. [40]

Deuterium wird als ausführbare Datei zur Verfügung gestellt und kann mit Hilfe des Kommandozeilenbefehls

$$\begin{aligned} &\text{DEUTERIUM} < \text{Summenformel} > < \text{Auflösung} > \\ &[-o < \text{Dateiname} >] [-L < \text{Isotopenzusammensetzung} >] [-p < \text{Masse} >] \quad (5.10) \\ &[-2D] [-v] [-d] \end{aligned}$$

gestartet werden. Folgende Kommandozeilenparameter müssen angegeben werden:

- *Summenformel*: Die Summenformel des Moleküls für die das Isotopenmuster berechnet werden soll. Sie wird in der Form *ElementAnzahl* - zum Beispiel *C378H629N105O118S* - angegeben.
- *Auflösung*: Das Spektrum wird mit der angegebenen Auflösung, wie zum Beispiel 100000, berechnet. Es handelt sich um die *FWHM* Definition der Auflösung.

Die folgenden Parameter sind optional:

- *-oDateiname*: Wird ein Dateiname angegeben, wird die Ausgabe in die entsprechende Textdatei abgespeichert. Es muss der vollständige Dateipfad enthalten sein.
- *-LIsotopenzusammensetzung*: Das Programm benötigt eine Datei in der die Massen und relativen Isotopenhäufigkeiten gespeichert sind. Wird keine Datei angegebenen wird *ELEMENTS.lib* als Standard verwendet. Die Datei muss im gleichen Ordner wie *Deuterium* enthalten sein. In dieser Arbeit wurde eine eigene Datei *ELEMENTS.lib* erstellt um sowohl für die multinomiale, als auch die 2D-FT Methode die gleichen Ausgangswerte zu schaffen (siehe Tabelle A-1).
- *-pPeak*: Mit dieser Option kann ein Peak an einer bestimmten Masse zurückgegeben werden. Die Berechnungszeit unterscheidet sich jedoch nicht von der Berechnungszeit wenn alle Peaks zurückgegeben werden. Diese Option wird ignoriert falls -2D gewählt ist.
- *-2D*: Mit dieser Option wird die Isotopenverteilung im 2D-Format zurückgegeben. Es besteht aus *n* Spalten die mit einem Tabulator (\t) voneinander getrennt sind. Jede Spalte enthält die Häufigkeiten der Peaks der Isotopenfeinstruktur an einer bestimmten nominellen Masse. Werden

³² Als Beispiel soll hier *C₃₉₄₈H₆₀₇₃N₁₁₂₃O₁₂₁₃S₅₉* genannt werden, welches mit einer Auflösung von $5 \cdot 10^8$ in zehn Sekunden berechnet wird. [37]

Auflösungen gewählt an der die Peaks überlappen wird nur dieses Format produziert. Es werden nur die Häufigkeiten ausgegeben, keine Massen.

- -d: Es werden nur die Ergebnisse ausgegeben und die Kopfzeilen unterdrückt.
- -v: Ausgabe zusätzlicher Informationen.

In dieser Arbeit wurde als Ausgabeformat die 1D-Darstellung verwendet, da hier die Massen und Häufigkeiten bereits in tabellarischer Form vorliegen. Es wurde der Befehl

DEUTERIUM < Summenformel > < Auflösung > -LMyElements.lib (5.11)

verwendet. In Abbildung 5-8 ist beispielhaft die Ausgabe des Befehls

DEUTERIUM C32H50N10I9 809793 - LMyElements.lib (5.12)

dargestellt. Die Zeilen enthalten der Reihe nach: Programmversion, Summenformel und Auflösung, Anzahl Peaks und Datenpunkte pro Peak, Berechnungsdauer und die Kopfzeilen der Tabelle. Ab Zeile sieben beginnen die Daten des Spektrums.

```
DEUTERIUM 0.1, Copyright (c) 2010 CIGB
C32H50N10I9 809793.000000
16      256
0.0150000000000000 seconds

mass      abundance
752.2975762084    0.9449876283
752.2977090000    1.0000000000
752.2978417916    0.9449876283
```

Abbildung 5-8: 1D Ausgabe von Deuterium

Es ist ein Ausschnitt der Ausgabe-Datei von der Berechnung des Spektrums von $C_{32}H_{50}NO_{19}$ dargestellt. Die erste Zeile enthält die Versionsbezeichnung von Deuterium. Die zweite Zeile die Summenformel und Auflösung. Die dritte Zeile die Anzahl an Peaks und Datenpunkte pro Peak. Die vierte Zeile die Berechnungszeit. Nach der Kopfzeile in Zeile 6 folgen die Daten des Spektrums.

Aufgrund der Verwendung der Fourier-Transformation und einer Peak-Form Funktion wird das simulierte Spektrum einer Summenformel als Profilspektrum berechnet. Für die Konvertierung in ein Linienspektrum werden die Maxima aller Peaks durch Differentiation bestimmt und zusammen mit den jeweiligen m/z – Werten abgespeichert. Die Intensitäten jeder simulierten Summenformel sind auf 1 normiert. Sie werden durch Multiplikation mit der Intensität des entsprechenden Peaks im

Originalspektrum angepasst. Da der 2D-FT Algorithmus die m/z -Werte nur für neutrale Moleküle berechnet, muss die Ladung von Ionen, die bei der FT-ICR-MS entstehen nach Formel 3.4 mit eingerechnet werden.

Aus dem Kommandozeilenbefehl Formel 5.11 geht hervor, dass bei der 2D-FT Methode eine Auflösung mit angegeben werden muss, um das Isotopenmuster zu berechnen. Das einzige Datenformat in dem die Auflösung mit angegeben ist, ist das TXT Format. Um auch bei anderen *Spektren* oder *Scans* ohne Daten über die Auflösung ein simuliertes Spektrum erstellen zu können, wird hier auf die eingestellte approximierte Auflösung zurückgegriffen.

5.3 Vergleich von Spektren

Das Programm SEAT beinhaltet eine Funktion um zwei *Spektren* oder *Scans*, unabhängig davon ob simuliert oder gemessen, miteinander zu vergleichen. Der Algorithmus basiert dabei auf der *Accurate Mass Spectrum Distance Methode (AMS)* [45].

5.3.1 Peak-Modell

Es sollen zwei Spektren Q und R paarweise verglichen werden. Der q -te bzw. r -te Peak eines (*simulierten*) *Spektrums* oder *Scans* wird mit Hilfe von m/z – Verhältnis, Intensität und Auflösung (*FWHM*) charakterisiert:

$$\begin{aligned}\phi_q^Q &= \{m_q^Q, i_q^Q, f_q^Q\} = \{m, i, f\}_q^Q, q = 1, \dots, |Q| \\ \phi_r^R &= \{m_r^R, i_r^R, f_r^R\} = \{m, i, f\}_r^R, r = 1, \dots, |R|\end{aligned}\tag{5.13}$$

Die Spektren Q und R werden somit durch $\Phi^Q = \{\phi_q^Q\}$ und $\Phi^R = \{\phi_r^R\}$ modelliert. In Abbildung 5-9 ist beispielhaft das Konzept des Peak-Modells dargestellt. Sind keine Daten über die Auflösung vorhanden, so wird für die jeweiligen m/z – Werte die eingestellte approximierte Auflösung berechnet.

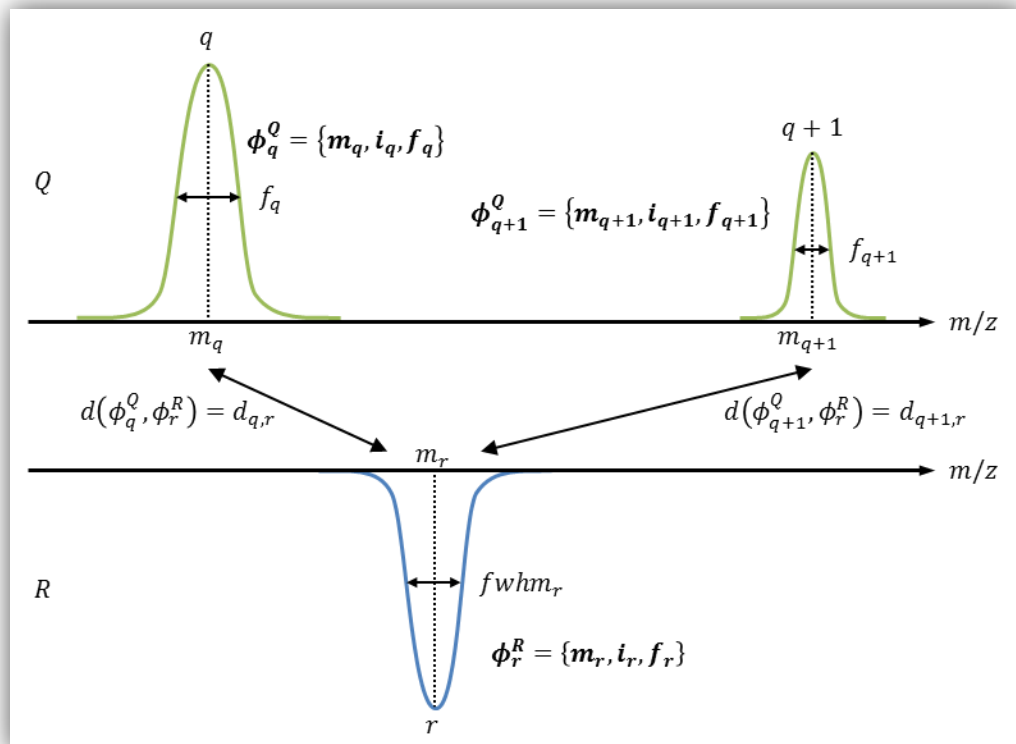


Abbildung 5-9: Peak-Modell und Distanz

Für alle Peaks q und r der zu vergleichenden Spektren Q und R werden Peak-Modelle ϕ erstellt, welche m/z -Werte (m), Intensität (i) und Auflösung (f) enthalten. Für alle Peak-Paarungen (ϕ_q^Q, ϕ_r^R) wird eine Distanz $d_{q,r}$ ermittelt, die die Ähnlichkeit der beiden Peaks zueinander widerspiegelt.

5.3.2 Begriffe für die Klassifikation von Peaks

Diese Peak-Modelle werden nun paarweise miteinander verglichen. Dabei wird jedes Peak-Paar (ϕ_q^Q, ϕ_r^R) anhand einer berechneten Distanz in eine der drei Kategorien eingeteilt:

Eine *Übereinstimmung* bzw. ein *Match* (M) liegt vor, wenn die Distanz des Peak-Paars kleiner ist als ein Schwellwert $t_{distance}$ und weder ϕ_q^Q noch ϕ_r^R schon einmal als Match klassifiziert worden ist. Die Kategorie *Keine Übereinstimmung* bzw. *Mismatch* (MM) enthält Peak-Paare, bei denen die Distanz ebenfalls kleiner als ein Schwellwert $t_{distance}$ ist, aber ϕ_q^Q oder ϕ_r^R schon einmal als Match klassifiziert wurde. Wenn ein Peak ϕ_q^Q bzw. ϕ_r^R weder in der Kategorie Match noch in der Kategorie Mismatch eingeteilt wurde und kein Peak-Paar (ϕ_q^Q, ϕ_z^R) , $z \neq r$ bzw. (ϕ_z^Q, ϕ_r^R) , $z \neq q$ mit einer Distanz kleiner dem Schwellwert t existiert, wird dieser als *Insertion* (I) bzw. *Deletion* (D) deklariert. In Abbildung 5-10 ist die Klassifikation der Peak – Paare grafisch dargestellt.

5.3.3 Distanz und Ähnlichkeit zwischen zwei Peaks

Auf der Grundlage von m/z und Auflösung

Als Maß für die Ähnlichkeit zwischen zwei beliebigen Peaks ϕ_q^Q und ϕ_r^R wird die Distanz $d(\phi_q^Q, \phi_r^R)$, kurz $d_{q,r}$, ermittelt. Vorteilhaft ist es eine Distanz zu wählen die sowohl den m/z -Wert als auch die Auflösung eines Peaks in die Berechnung einbezieht um auch sehr dicht liegende Peaks gut klassifizieren zu können. Betrachtet man zum Beispiel die Peak-Paare (C, F) und (C, G) in Abbildung 5-10 so ist es schwierig nur anhand des m/z -Wertes eine Entscheidung zu treffen, welches der beiden Paare als Match eingestuft werden soll. Wird die Auflösung in die Bewertung mit einbezogen, ist deutlich, dass (C, F) ähnlicher zueinander ist als (C, G) . Des Weiteren könnte man die Intensität für die Bewertung der Ähnlichkeit von Peaks hinzufügen. Jedoch können die Intensitäten der einzelnen Peaks, auch bei genormten Spektren³³ stark voneinander abweichen, wie zum Beispiel bei Spektren von Kollisionsexperimenten, bei denen der Basispeak zu unterschiedlichen Ionen zugeordnet sein kann.

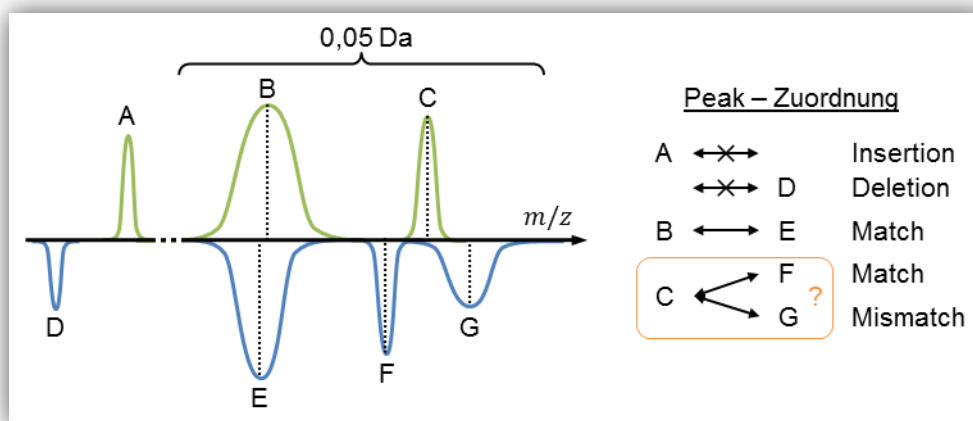


Abbildung 5-10: Klassifikation der Peaks

Jedes Peak-Paar bzw. jeder Peak wird, abhängig von der Distanz, in eine von drei Kategorien eingeteilt. **Links** ist ein Ausschnitt zweier Spektren dargestellt. **Rechts:** Die Klassifikation der Peak-Paare. Unter der Annahme, dass die Distanz von dem Peak-Paar (B, E) kleiner als ein Schwellwert ist, gilt dieses Paar als Match. Ebenso bei (C, F) . Da (C, G) ebenfalls eine geringere Distanz besitzt, doch Peak C schon einmal in der Kategorie Match gelistet ist, wird dieses Peak-Paar als Mismatch klassifiziert.

³³ Es ist die Normierung des Spektrums auf relative Intensität anhand des Basispeaks gemeint.

Als eine mögliche Distanzmetrik wurde die *Jeffrey-Matusita Distanz*³⁴

$$d_{q,r} = JM_{q,r} = \frac{1}{2} \left(\int_m \left(\sqrt{g(m, \phi_q)} - \sqrt{g(m, \phi_r)} \right)^2 dm \right)^{1/2} \quad (5.14)$$

verwendet. Für diese Gleichung wird eine Peak-Funktion $g(m, \phi_\bullet)$ ³⁵ für die Peak-Modelle ϕ_\bullet benötigt. Wie schon in Kapitel 2.3.5 erläutert, stellt die Gauß-Verteilung eine gute Approximation der Peaks, bzw. der natürlichen Ionenpopulation, dar. Somit ist

$$g(m, \phi_\bullet) = i_\bullet \cdot e^{\frac{(m-m_\bullet)^2}{2\sigma_\bullet^2}}. \quad (5.15)$$

Durch die Verwendung der Gauß-Verteilung vereinfacht sich Formel 5.14 zu

$$d_{q,r} = JM_{q,r} = \sqrt{(1 - e^{-\alpha_{q,r}})}, \quad (5.16)$$

wobei $\alpha_{q,r}$ die *Bhattacharyya-Distanz* darstellt. Sie berechnet sich durch

$$\alpha_{q,r} = \frac{1}{4} \frac{(m_q - m_r)^2}{\sigma_q^2 + \sigma_r^2} + \frac{1}{2} \ln \left(\frac{\sigma_q^2 + \sigma_r^2}{2 \cdot \sigma_q \cdot \sigma_r} \right). \quad (5.17)$$

Die Variablen m_q und m_r sind die jeweiligen m/z – Werte der Peak-Modelle ϕ_q bzw. ϕ_r und σ_q bzw. σ_r die Standardabweichungen der Gauß-Verteilung nach Formel 2.22. Die Bhattacharyya Distanz besteht aus zwei Summanden. Der erste Summand beschreibt die Distanz zwischen den m/z – Werten (Mittelwert der Verteilung) und der Zweite Summand berechnet die Distanz in Bezug auf die Auflösung eines Peaks (Varianz der Verteilung). Es gilt $\alpha_{q,r} = \alpha_{r,q}$ und somit $JM_{q,r} = JM_{r,q}$. **[46, 47, 45, 48]**

Der Wertebereich der Bhattacharyya Distanz ist $\alpha \in [0, \infty)$. Damit ergibt sich für die Jeffrey-Matusita Distanz ein Wertebereich von $JM \in [0, 1]$, wobei zwei Peaks mit einer Distanz von 0 identisch und mit einer Distanz von 1 komplett verschieden sind.

³⁴ Die Jeffrey-Matusita Distanz wird in der Literatur auch als Hellinger Distanz oder Matusita Distanz bezeichnet. [52]

³⁵ Das Zeichen • wird als Platzhalter für q oder r verwendet.

Auf der Grundlage des ppm-Werts

Ist ein Vergleich auf Basis des m/z -Wertes und der Auflösung nicht erwünscht oder nicht möglich, kann auch ein einfacher Vergleich der Peaks mit Hilfe des ppm-Wertes erfolgen. Dafür wird folgende Distanz gewählt:

$$d_{q,r} = \begin{cases} \left\lceil \frac{ppm_{q,r}}{t_{ppm}} \right\rceil & \text{wenn } ppm_{q,r} \leq t_{ppm}, ppm_{q,r} = 10^6 \cdot \frac{(m/z)_q - (m/z)_r}{((m/z)_q + (m/z)_r)/2} \\ 1 & \text{sonst} \end{cases} \quad (5.18)$$

Der Wert t_{ppm} stellt den Schwellwert, bis zu dem ein Peak-Paar als Match klassifiziert werden soll, dar.

5.3.4 Distanzmatrix und Klassifikation der Peak-Paare

Ziel des Algorithmus ist es nun, mit Hilfe der Distanz $d_{q,r}$ korrespondierende Peak-Paare der Spektren Q und R zu finden. Nach der Erstellung aller Peak-Modelle ϕ , werden die Distanzen aller Peak-Paare (ϕ_q^Q, ϕ_r^R) nach Formel 5.16 und Formel 5.17 bzw. Formel 5.18 berechnet. Daraus resultiert die Distanzmatrix \mathbb{D} mit

$$\mathbb{D} = \begin{pmatrix} d_{1,1} & \cdots & d_{1,|R|} \\ \vdots & \ddots & \vdots \\ d_{|Q|,1} & \cdots & d_{|Q|,|R|} \end{pmatrix} \text{ mit } \mathbb{D} \in \mathcal{R}^{|U| \times |R|}. \quad (5.19)$$

Die Peak-Paare werden nun anhand ihrer Distanz, beginnend mit der kleinsten, sortiert. Des Weiteren wird ein Schwellwert $d_t \in (0,1)$ festgelegt. [45] Für die Klassifizierung der Peaks werden die unter Kapitel 5.3.2 definierten Kategorien verwendet und es werden folgende Bedingungen festgelegt:

1. Ein Peak ϕ_q^Q kann nur einem einzigen Peak ϕ_r^R zugeordnet werden und andersherum. Es sind keine Mehrfachzuordnungen zulässig.
2. Die Peaks dürfen in ihrer Reihenfolge nicht vertauscht werden. Beispielsweise ist die Zuordnung $\phi_q^Q \leftrightarrow \phi_{r+1}^R$ und $\phi_{q+1}^Q \leftrightarrow \phi_r^R$ nicht zulässig.

Die sortierte Peak-Paar Liste Λ wird nun verwendet um die einzelnen Peak-Paare nach Formel 5.20 zu klassifizieren. Dabei wird jedes Peak-Paar, je nach Kategorie, mit einem Index $K \in \{M, MM, I, D\}$ ³⁶ versehen.

³⁶ Die Indices stehen für M=Match, MM=Mismatch, D=Deletion, I=Insertion.

$$(\phi_q^Q, \phi_r^R)_K = \begin{cases} M & \text{wenn } (d_{q,r} \leq d_t) \wedge (\phi_q^Q \notin \Lambda_M) \wedge (\phi_r^R \notin \Lambda_M) \\ MM & \text{wenn } (d_{q,r} \leq d_t) \wedge ((\phi_q^Q \in \Lambda_M) \vee (\phi_r^R \in \Lambda_M)) \\ I & \text{wenn } \forall \phi_q^Q: d_{q,r} > d_t \text{ mit } r = 1, \dots, |R| \\ D & \text{wenn } \forall \phi_r^R: d_{q,r} > d_t \text{ mit } q = 1, \dots, |Q| \end{cases} \quad (5.20)$$

Damit besitzt die Peak-Paar-Liste Λ vier Teillisten $\Lambda_M, \Lambda_{MM}, \Lambda_D, \Lambda_I$ mit $\Lambda_M \cup \Lambda_{MM} \cup \Lambda_D \cup \Lambda_I = \Lambda$ und $\Lambda_M \cap \Lambda_{MM} \cap \Lambda_D \cap \Lambda_I = \emptyset$. Da es sich bei den Deletionen bzw. Insertionen nicht um Peak-Paare handelt, sondern die Peaks einzeln betrachtet werden, wird die Schreibweise $(\phi_q^Q, 0)_I$ oder $(\phi_q^Q)_I$ bzw. $(0, \phi_r^R)_D$ oder $(\phi_r^R)_D$ verwendet.

5.3.5 Ähnlichkeit zwischen Spektren

Die paarweise Klassifikation gibt Aufschluss über die Ähnlichkeit zwischen zwei Peaks. Um eine Aussage über die Ähnlichkeit zwischen zwei Spektren treffen zu können, wird ein weiteres Distanz-Maß, welches mit Hilfe der Liste Λ_M berechnet wird, benötigt. Es werden alle Distanzen eines Peaks-Paares $(\phi_q, \phi_r)_M$ summiert und mit einem Faktor ω_0 , der den Mittelwert aus den Verhältnissen von gepaarten zu nicht gepaarten Peaks der beiden Spektren Q und R widerspiegelt, multipliziert. Die Gesamt-Distanz $d_{Q,R}$ berechnet sich schließlich mit der Formel

$$d_{Q,R} = \frac{\omega_0}{L} \cdot \sum_{l=1}^L (d_{q,r})_l \text{ mit } \omega_0 = \frac{\omega_Q + \omega_R}{2} \quad (5.21)$$

wobei

$$\omega_Q = 1 - \frac{L}{|Q|} \text{ und } \omega_R = 1 - \frac{L}{|R|} \quad (5.22)$$

und L die Anzahl an gefundenen Matches. [45]

6 Beispielberechnungen mit SEAT

Die Funktionalität des Programms wurde mit Hilfe der SAIB-Spektren getestet. Hier werden kurz die Ergebnisse der Simulation und des Vergleichs von Spektren vorgestellt. Des Weiteren werden die Ergebnisse einiger Laufzeitmessungen präsentiert. Getestet wurde SEAT auf einem Windows 7 64-bit Betriebssystem mit 8 GB RAM und einem Intel Core i7 bei 1,6 GHz.

6.1 Simulation mit der multinomialen Methode

6.1.1 SAIB Spektren

Von dem fragmentierten und nicht fragmentierten Spektrum von SAIB wurden jeweils sechs Spektren mit den Schwellwerten 10^{-1} , 10^{-2} , ..., 10^{-7} und einer Ladung von -1 berechnet. In Abbildung 6-1 ist beispielhaft ein Ausschnitt des nicht fragmentierten Spektrums (positive Achse, grün) und dem entsprechenden simulierten Spektrum mit Schwellwert 10^{-3} (negative Achse, blau) abgebildet. Zur besseren Übersicht wurde auf die Anzeige der Summenformeln verzichtet, jedoch wurden alle Peaks, zu denen Summenformeln zugeordnet sind, im nicht fragmentierten Spektrum markiert (magentafarbene Peaks).

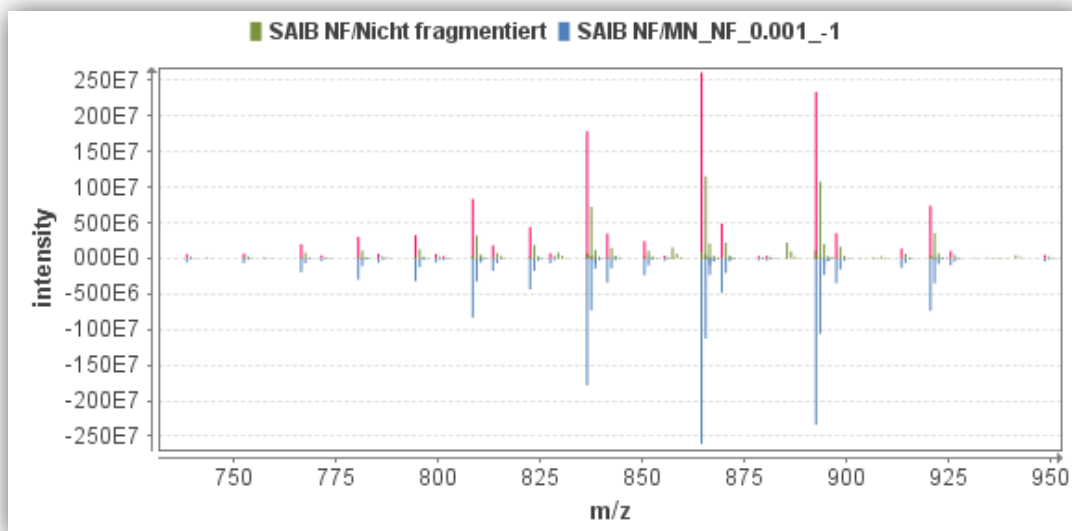


Abbildung 6-1: Nicht fragmentiertes Spektrum und multinomial simuliertes Spektrum im Vergleich
Auf der positiven Achse ist das nicht fragmentierte Spektrum abgebildet. Die in Magenta hervorgehobenen Peaks sind mit einer Summenformel assoziiert. Das blaue Spektrum auf der negativen Achse ist das mit der multinomialen Methode simulierte Spektrum mit einem Schwellwert von 0,001 und Ladung von -1.

Alle simulierten Spektren ergeben ein ähnliches Bild. Unterschiedlich ist jedoch die Anzahl der berechneten Peaks pro Summenformel, also die Anzahl der Isotopenpeaks, abhängig von dem gewählten Schwellwert. Sie reicht von zwei Peaks bei einem Schwellwert von 10^{-1} bis hin zu 116 Peaks bei einem Schwellwert von 10^{-7} und steigt im Allgemeinen bei sinkendem Schwellwert. Damit steigt auch die Gesamtzahl der Peaks im simulierten Spektrum. Die benötigte Rechenzeit liegt bei allen Schwellwerten und beiden Spektren zwischen 25 ms und 129 ms und damit unter einer Sekunde. Tabelle 6-1 fasst die Ergebnisse kurz zusammen.

Tabelle 6-1: Berechnung der SAIB Spektren mit der multinomialen Methode

Die Tabelle listet die Laufzeit, die Gesamtzahl der Peaks sowie die minimale und maximale Anzahl von Peaks mit der entsprechenden Summenformel auf. Weiße Zeilen entsprechen dem fragmentierten Spektrum (F) und Blaue dem nicht fragmentierten (NF)

Schwellwert		10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	Bemerkung
Laufzeit in [s]	NF	0,058	0,025	0,025	0,031	0,042	0,053	0,065	
	F	0,083	0,039	0,051	0,058	0,081	0,107	0,129	
Gesamtzahl Peaks	NF	60	151	344	626	1064	1665	2384	Original:391
	F	149	346	761	1363	2275	3510	5011	Original:1763
Minimale Anzahl Peaks	NF	2	5	13	23	40	66	95	
	F	2	2	6	10	13	23	29	
Maximale Anzahl Peaks	NF	3	6	15	28	49	76	114	
	F	3	6	15	29	49	78	116	

In Abbildung 6-2 ist die Anzahl der Peaks pro Isotopenmuster (Summenformel) des nicht fragmentierten Spektrums und in Abbildung A-1 des fragmentierten Spektrums für alle Schwellwerte dargestellt. Die Summenformeln sind nach der Masse aufsteigend sortiert. Beide zeigen einen ähnlichen Verlauf. Mit zunehmender Masse und sinkendem Schwellwert werden mehr Peaks berechnet. Des Weiteren gibt es relativ starke Schwankungen in der Peak-Anzahl bei gleichbleibendem Schwellwert. Diese sind durch die unterschiedliche atomare Zusammensetzung der Summenformeln zu erklären. Es gibt Summenformeln die neben Wasserstoff, Sauerstoff und Kohlenstoff entweder ein Natrium- ein Stickstoffatom oder keines von beiden enthalten. Da Stickstoff zwei Isotope besitzt und Natrium nur eines, wirkt sich dies insgesamt auf die Anzahl möglicher Isotopenkombinationen und deren Häufigkeiten aus. Dies wiederum führt zu den Schwankungen in der Peak-Anzahl. Unabhängig von der elementaren Zusammensetzung ist die steigende Peak-Anzahl bei zunehmender Masse ebenfalls durch die zunehmende Anzahl möglicher Isotopenkombinationen erklärbar. Je mehr

Atome in einem Molekül vorhanden sind, desto höher ist die Wahrscheinlichkeit, dass seltenere Isotope in diesem Molekül vorkommen.

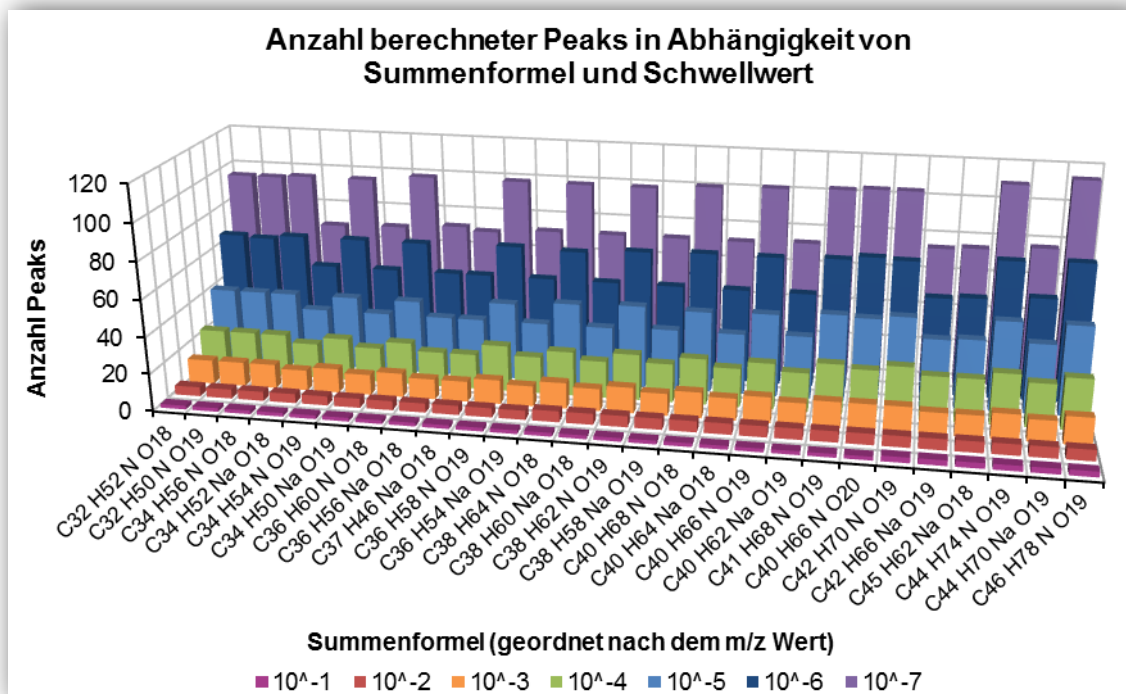


Abbildung 6-2: Anzahl berechneter Peaks in Abhängigkeit von Summenformel und Schwellwert

Die Anzahl der Peaks pro Summenformel nimmt mit steigender Masse und sinkendem Schwellwert zu.

Die starken Schwankungen der Peak-Anzahl bei gleichbleibendem Schwellwert sind durch unterschiedliche Isotopenzahlen der Elemente erklärbar.

6.1.2 Laufzeitmessungen

Für die Einschätzung der Laufzeit und Effizienz des multinomialen Algorithmus wurden die Berechnungsdauer in Abhängigkeit von der Anzahl der Atome und Isotope der Elemente untersucht. Es wurden Summenformeln bestehend aus einem Element mit zwei bis sieben Isotopen³⁷ und $n = 10, 20, \dots, 100$ Atomen sowie einem Schwellwert von 0,001 getestet. Das Ergebnis ist in Abbildung 6-3 dargestellt. Die Laufzeiten sind logarithmisch aufgetragen. Man erkennt, dass die Berechnungsdauer mit Ausnahme von Kohlenstoff bei zunehmender Atomanzahl linear bei drei und vier Isotopen zunimmt. Dies entspricht einem exponentiellen Wachstum bei nicht logarithmischer Skalierung. Ab fünf Isotopen wächst die Berechnungsdauer faktoriell. Des Weiteren nimmt die Laufzeit bei steigender Isotop-Anzahl und gleichbleibender Atomanzahl zu. Eine Erklärung dafür ist, dass je mehr Atome ein Molekül und je mehr Isotope das

³⁷ Gewählt wurden als Elemente Kohlenstoff (C), Sauerstoff (O), Schwefel (S), Zink (Zn), Calcium (Ca) und Quecksilber (Hg)

entsprechende Element besitzt, desto mehr Isotopenkombinationen treten auf. Zum Beispiel gibt es für O_{50} insgesamt 1.326 und für Ca_{50} schon 3.478.761 mögliche Isotopenkombinationen. Aufgrund der hohen Berechnungsdauer sind für die Isotopenzahlen fünf, sechs und sieben nicht alle Werte bis $n = 100$ verfügbar

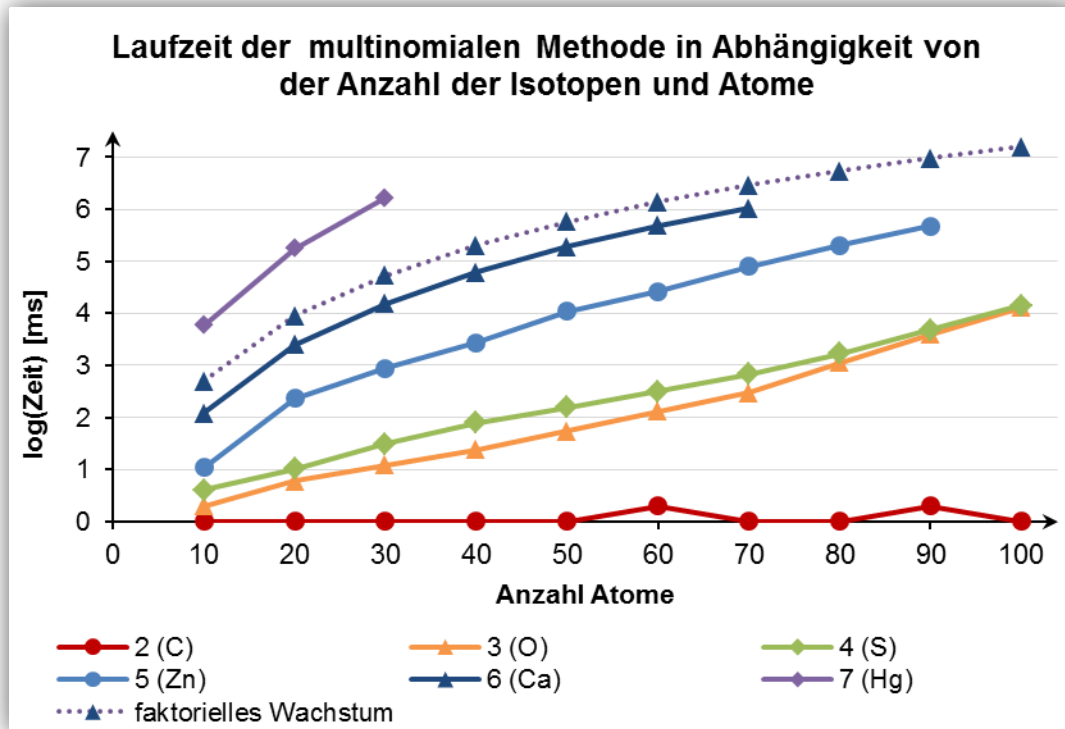


Abbildung 6-3: Laufzeit der multinomialen Methode in Abhängigkeit von Isotopen- und Atomanzahl

Es ist die logarithmische Laufzeit der multinomialen Methode für verschiedene Elemente mit zwei bis sieben Isotopen dargestellt. Die benutzten Elemente sind in Klammern hinter den Isotopenzahlen in der Legende angezeigt. Es wurden Summenformeln bestehend aus zehn bis maximal 100 Atomen eines Elements für die Analyse verwendet. Die verwendete Auflösung beträgt 10^7 . Die gepunktete Linie entspricht der abgeschätzten faktoriell wachsenden Laufzeit für sechs Isotope.

Um die Laufzeit von realen Summenformeln abschätzen zu können wurden eine Test-Datei mit Summenformeln mit fünf unterschiedlichen elementare Zusammensetzungen gewählt. Da das Programm für die Auswertung von Pyrolyseöl-Spektren entwickelt wurde, enthalten alle Summenformeln Kohlenstoff, Wasserstoff und Sauerstoff. Insgesamt wurden die fünf theoretischen Summenformeltypen

$$\begin{aligned}
 &C_n H_{2n+2}, \quad C_n H_{2n+2} O_n, \quad C_n H_{2n+2+m} O_n N_m, \\
 &C_n H_{2n+2+m} O_n N_m S_m, \quad C_n H_{2n+2+m} O_n N_m S_m F e_{m/2}
 \end{aligned}
 \tag{6.1}$$

mit $n = 10, 20, \dots, 500$ und $m = 2, 4, \dots, 100$ verwendet. Es sind Atome mit zwei (Kohlenstoff, Wasserstoff und Stickstoff), drei (Sauerstoff) und vier Isotopen (Schwefel

und Eisen) vertreten. In Abbildung 6-6 ist die benötigte Rechenzeit logarithmisch für die fünf Summenformeltypen dargestellt. Aufgrund der schnell wachsenden Berechnungsdauer pro Isotopen-Anzahl und Atomanzahl zeigt das Diagramm die Ergebnisse nur bis $n = 130$, $m = 26$. Man erkennt einen linearen, bzw. bei nicht logarithmischer Darstellung exponentiellen Zusammenhang zwischen der Anzahl der Atome und der Laufzeit. Zwischen den einzelnen Summenformeltypen ist die Laufzeit annähernd konstant. Eine Ausnahme bildet hier die der Typ $C_n H_{2n+2}$, bei dem die Laufzeit durchgehend konstant bleibt. Dies ist damit zu begründen, dass Kohlenstoff und Wasserstoff diisotopische Elemente sind und die Isotopenverteilungen nicht mit Hilfe des Partitionsalgorithmus berechnet werden müssen (vgl. Kapitel 5.2.1).

Für die Berechnung des Isotopenmusters der Summenformeln mit kleinster Masse $C_{10}H_{22}$ benötigt der Algorithmus weniger als 10 ms und bei der Summenformel mit der größten Masse $C_{130}H_{288}O_{120}N_{26}S_{26}Fe_{13}$ 173 Sekunden.

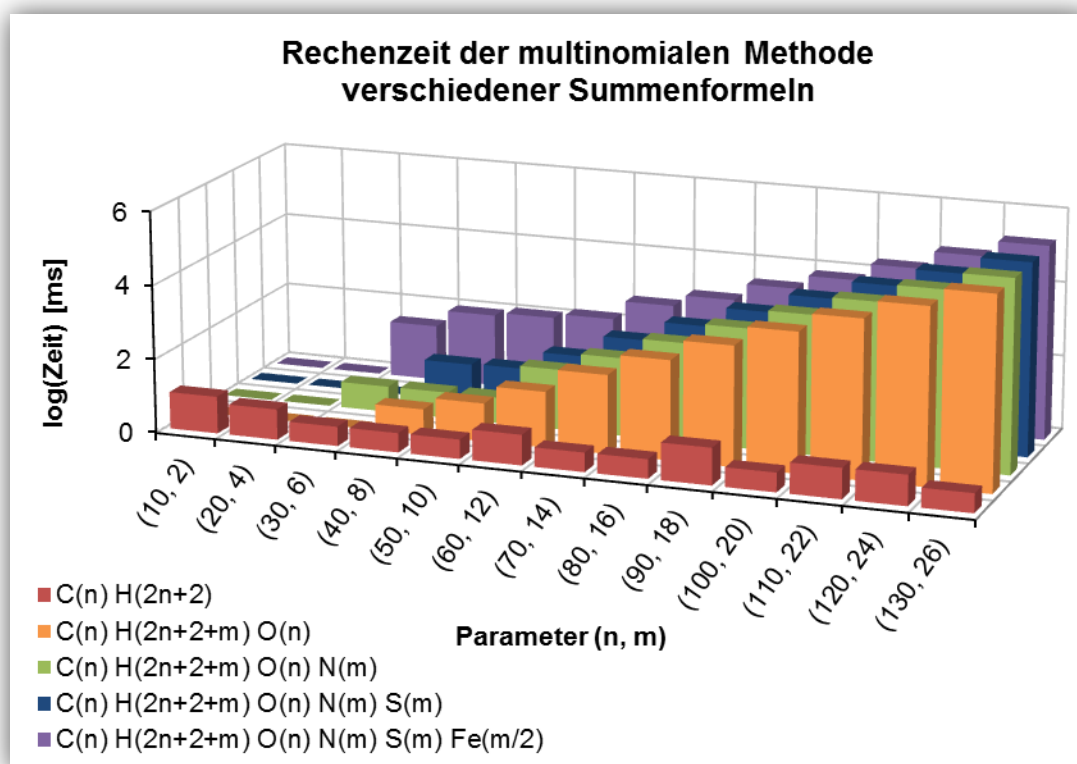


Abbildung 6-4: Rechenzeit der multinomialen Methode von fünf Summenformeltypen

Es ist die Rechenzeit verschiedener Summenformeln dargestellt. Die Rechenzeit ist logarithmisch aufgetragen und in Millisekunden angegeben. Es ist ein linearer Zusammenhang zwischen der Anzahl der Atome der Elemente in den Summenformeln und der Rechenzeit erkennbar. Eine Ausnahme bildet $C_n H_{2n+2}$, da hier nur diisotopische Elemente vorhanden sind. Die Anzahl der Elemente in den Summenformeln hat wenig bis kaum Einfluss auf die Rechenzeit.

6.2 Simulationen mit der 2D-FT Methode

6.2.1 SAIB Spektren

Für die Simulation des fragmentierten und nicht fragmentierten Spektrums mit der 2D-FT Methode wurde die Ladung -1 und fünf verschiedene Auflösungen verwendet. Zum einen die gemessenen Auflösungen aus den zugehörigen TXT Dateien sowie die nach Formel 5.4 und den Standardwerten approximierte Auflösung und zum anderen die konstanten Werte 10^4 , 10^5 , 10^6 und 10^7 . Das Ergebnis der Simulation mit dem 2D-FT Algorithmus mit der gemessenen Auflösung ist in Abbildung 6-5 dargestellt. Auf der positiven Achse ist das nicht fragmentierte Spektrum und auf der negativen Achse das simulierte Spektrum abgebildet. Die Peaks, zu denen Summenformeln gehören wurden im nicht fragmentierten Spektrum markiert.

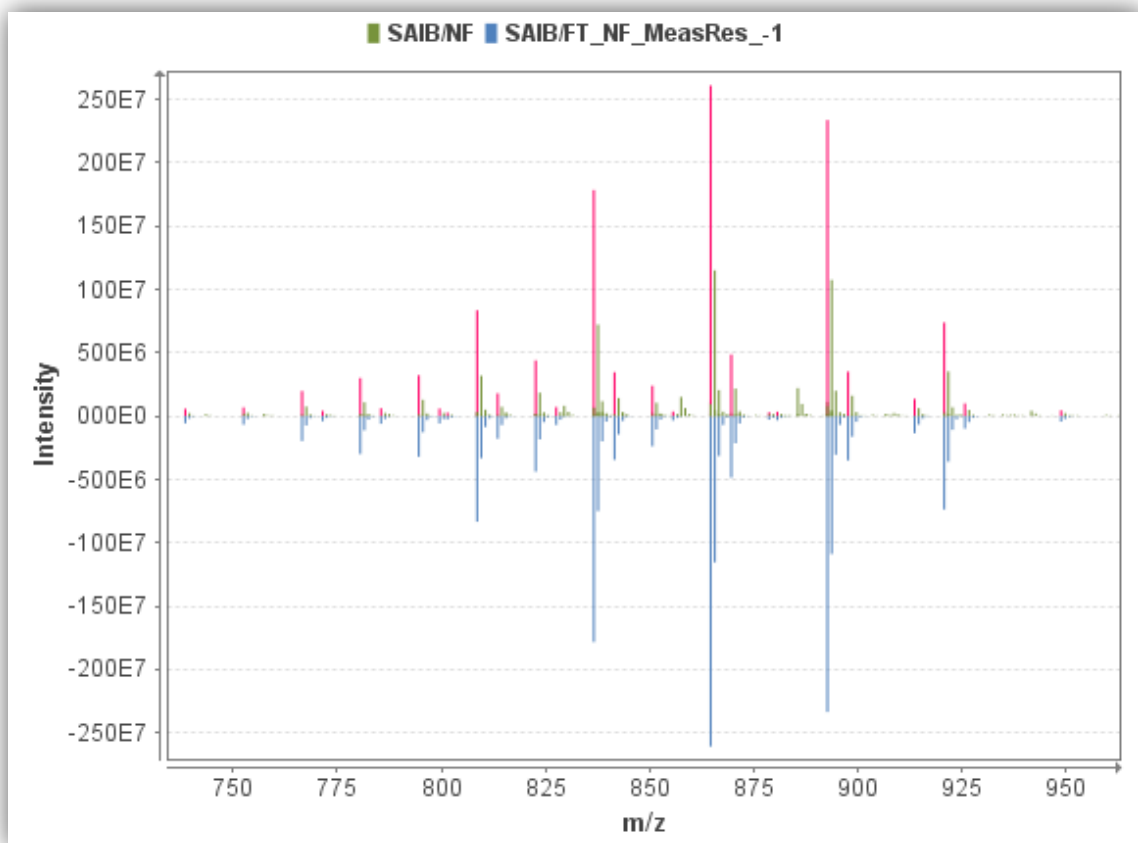


Abbildung 6-5: Nicht fragmentiertes Spektrum & 2D-FT simuliertes Spektrum im Vergleich

Auf der positiven Achse ist das nicht fragmentierte Spektrum abgebildet. Die in Magenta hervorgehobenen Peaks sind mit einer Summenformel assoziiert. Das blaue Spektrum auf der negativen Achse ist das mit der multinomialen Methode simulierte Spektrum mit der gemessenen Auflösung und Ladung von -1 .

Wie schon bei der multinomialen Methode unterscheiden sich die simulierten Spektren in der Anzahl simulierter Peaks pro Summenformel. Mit steigender Auflösung nimmt die Anzahl der Peaks zu. Im Allgemeinen ist die Anzahl der berechneten Peaks mit der 2D-FT Methode jedoch höher als bei der multinomialen. Die Laufzeit ist im Vergleich zu der multinomialen Methode mit 1,3 s beim nicht fragmentierten bzw. 3,4 s beim fragmentierten bei der Benutzung der gemessenen Auflösung bis zu 7,9 s bzw. 28,47 s bei einer Auflösung von 10^7 deutlich höher als bei der multinomialen Methode. Je höher die Auflösung desto länger dauert die Berechnung und desto mehr Peaks werden berechnet. In Tabelle 6-2 sind die Laufzeit und Peak-Anzahl kurz zusammengefasst.

Tabelle 6-2: Ergebnisse der Berechnung der SAIB Spektren mit der 2D-FT Methode

Die Tabelle listet die Laufzeit, die Gesamtzahl der Peaks sowie die minimale und maximale Anzahl von Peaks mit der entsprechenden Summenformel auf. Weiße Zeilen entsprechen dem fragmentierten Spektrum (F) und Blaue dem nicht fragmentierten (NF). MEAS steht für die Verwendung der gemessenen Auflösung und EXP2 für die approximierte Auflösung nach Formel 5.4.

Auflösung		MEAS	EXP2	10^4	10^5	10^6	10^7	Bemerkung
Laufzeit in [s]	NF	1,308	1,361	1,207	1,246	1,648	7,936	
	F	3,407	3,341	3,004	3,796	4,280	28,478	
Anzahl Peaks	NF	244	237	229	272	1228	3936	Original:391
	F	1025	979	565	640	3226	8442	Original:1763
Minimale	NF	8	8	8	8	35	102	
Anzahl Peaks	F	8	8	8	8	32	55	
Maximale	NF	14	13	9	12	65	191	
Anzahl Peaks	F	48	49	9	15	70	195	

Die Anzahl der Peaks für das nicht fragmentierte Spektrum ist in Abbildung 6-6 und für das fragmentierte Spektrum in Abbildung A-2 dargestellt. Bei beiden Spektren liegt die Anzahl der Peaks für die gemessenen, approximierten und konstanten ($10^4, 10^5$) Auflösungen zwischen 8 und 12 Peaks. Im fragmentierten Spektrum bei der gemessenen und approximierten Auflösung erkennt man zudem, dass bei einigen Summenformeln mit kleinerer Masse bis zu 60 Peaks berechnet worden sind. Dies ist Aufgrund der exponentiell fallenden Auflösungsfunktion erklärbar. Je geringer die Masse einer Summenformel, desto höher die Auflösung und desto mehr Peaks können voneinander getrennt gemessen werden. Da der 2D-FT- Algorithmus die Isotopenmuster auf Basis der Auflösung berechnet, gilt dieser Zusammenhang ebenfalls.

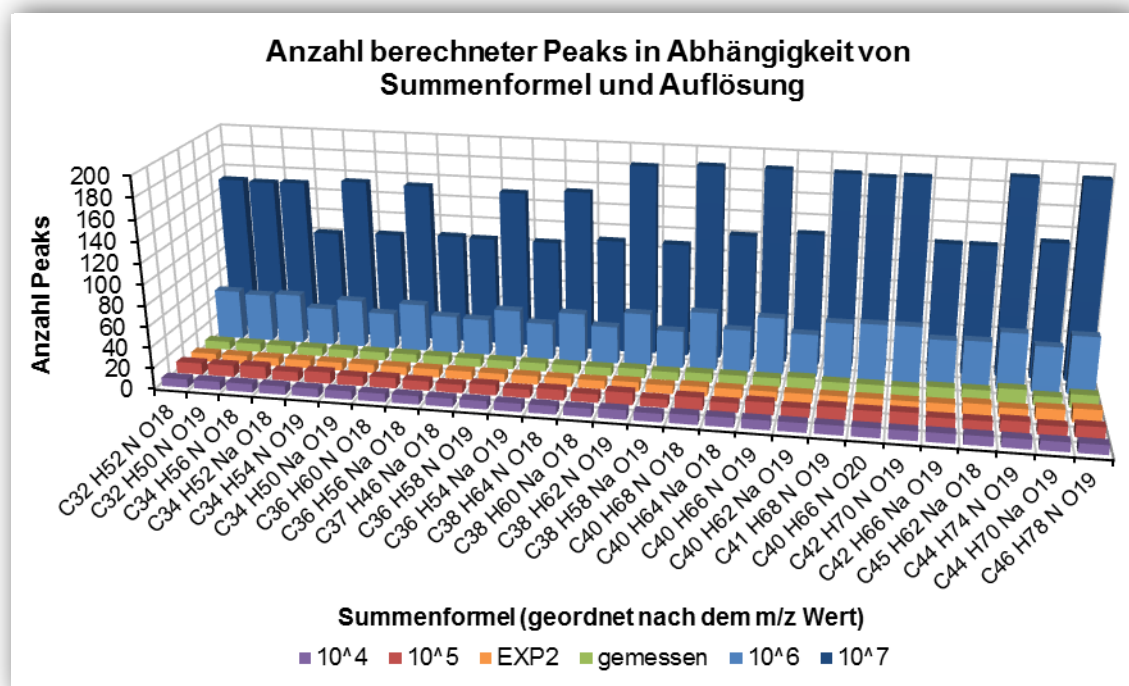


Abbildung 6-6: Anzahl berechneter Peaks in Abhängigkeit von Summenformel und Auflösung

Die Anzahl der Peaks pro Summenformel nimmt mit zunehmender Masse und Auflösung zu. Die starken Schwankungen der Peak-Anzahl bei gleichbleibendem Schwellwert sind durch unterschiedliche Isotopenzahlen der Elemente erklärbar.

6.2.2 Laufzeitmessungen

Für den 2D-FT Algorithmus wurden ebenfalls einige Laufzeitmessungen durchgeführt. Es wurden die Isotopenmuster für theoretische Moleküle bestehend aus einem Element mit $n = 10, \dots, 100$ Atomen berechnet und die Laufzeit gemessen. Die Auflösung betrug 10^7 . Abbildung 6-7 zeigt das Ergebnis in Abhängigkeit von Isotopen- und Atom-Anzahl. Anders als bei der multinomialen Methode bewegt sich die Laufzeit bei allen Elementen innerhalb eines Zeitintervalls von minimal 50 ms bei zwei Isotopen bis maximal 30 Sekunden bei neun Isotopen. Innerhalb eines Isotops schwanken die Werte ca. um den Faktor 10 und ist für jedes Element unterschiedlich. Dies ist durch die unterschiedlichen Isotopenhäufigkeiten und den unterschiedlichen Massen der Isotope erklärbar. Zudem ist die Laufzeit von der eingesetzten Auflösung abhängig. Da die 2D-FT Methode nicht wie die multinomiale Methode jede Isotopenkombination einzeln berechnet sondern viel mehr die Summe aller Kombinationen, ist es abhängig von der Auflösung, Masse und Häufigkeit ob zwei Isotopenverteilungen als getrennte oder als aggregierter Peak im berechneten Spektrum erscheint. Dies wiederum beeinflusst die Anzahl der Peaks und damit die Berechnungsdauer.

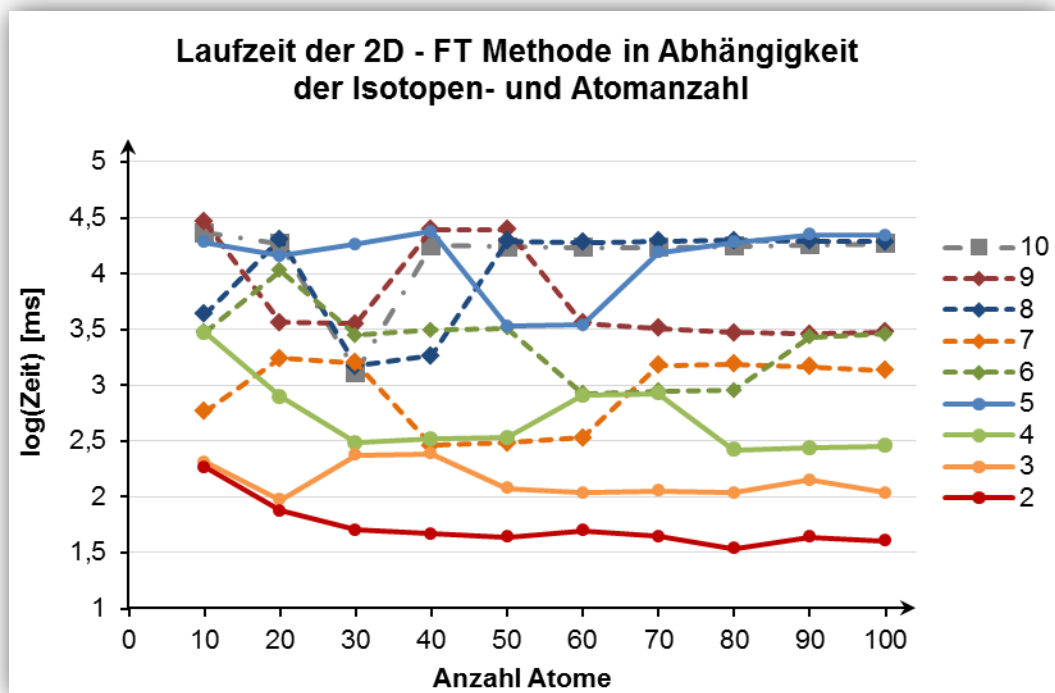


Abbildung 6-7: Laufzeit der 2D-FT Methode in Abhängigkeit von Isotopen- und Atomanzahl

Es ist die logarithmische Laufzeit in Millisekunden der 2D-FT Methode für verschiedene Elemente mit zwei bis zehn Isotopen dargestellt. Die benutzten Elemente sind in Klammern hinter den Isotopenzahlen in der Legende angezeigt. Es wurden Summenformeln bestehend aus zehn bis maximal 100 Atomen eines Elements für die Analyse verwendet.

Für die 2D-FT Methode wurden ebenfalls Laufzeitmessungen für die fünf Summenformeltypen aus Formel 6.1 vorgenommen. Es wurde eine Auflösung von 10^7 verwendet. Die logarithmische Laufzeit in Millisekunden in Abhängigkeit der Summenformeln ist in Abbildung 6-8 dargestellt. Es ist auch hier eine konstante Laufzeit bei steigender Atomanzahl innerhalb eines Summenformeltyps erkennbar. Die geringste Laufzeit besitzt C_nH_{2n+2} mit 80 ms bis 1 s. Die höchste Laufzeit hat $C_nH_{2n+2+m}O_nN_mS_mFe_{m/2}$ mit 1,5 s bis 14,2 s. Im Anhang in Tabelle A-2 sind die Laufzeiten der beiden Methoden zu Simulation von Spektren vergleichend gegenübergestellt.

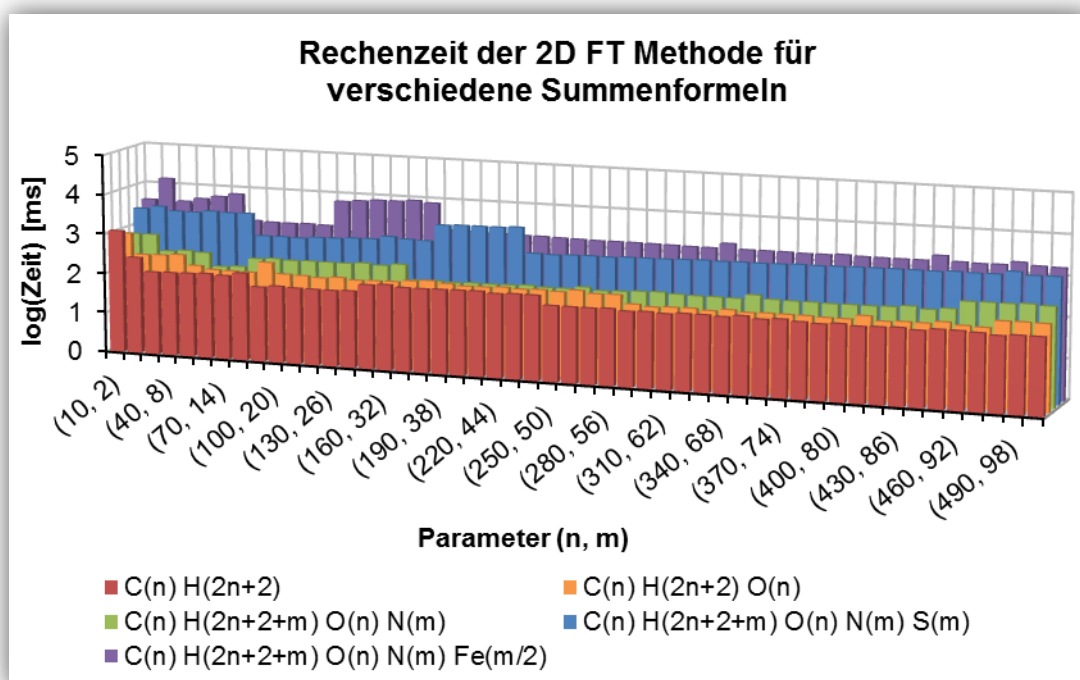


Abbildung 6-8: Rechenzeit der 2D – FT Methode für verschiedene Summenformeln

Es ist die logarithmische Rechenzeit in Millisekunden verschiedener Summenformeltypen logarithmisch aufgetragen. Sie bleibt bei zunehmender Anzahl an Elementen und Atomen pro Element annähernd konstant.

6.3 Vergleich von Spektren

6.3.1 Anwendung auf gemessene und simulierte Spektren

Der Vergleich von Spektren ermöglicht es zum Beispiel die Isotopenpeaks aus einem Spektrum mit Hilfe eines simuliertes Spektrums zu filtern, kann dabei helfen noch nicht erkannte Isotopenmuster zu finden oder Seitenpeaks bei sehr starken Intensitäten von Isotopenpeaks zu unterscheiden. Es wurden verschiedene simulierte Spektren mit gemessenen und die gemessenen Spektren untereinander verglichen. In Tabelle 6-3 sind die Ergebnisse einiger Vergleiche dargestellt. Für den Distanz-Schwellwert, der entscheidet ob ein Peak als Match klassifiziert wird, wurde für die Verwendung der Jeffrey-Matusita Distanz $t_{distance} = 0.5$ gewählt. Für den ppm-Schwellwert wurde $t_{ppm} = 0.5$ und $t_{distance} = 1$ gewählt. Das bedeutet alle Peak-Paare mit einem ppm-Wert kleiner 0.5 werden als Match eingestuft, insofern nicht einer der beiden Peaks vorher schon einmal als Match eingestuft worden ist.

Die Ergebnisse zeigen, dass bei einem Vergleich der gemessenen Spektren miteinander mehr übereinstimmende Peaks gefunden worden als bei dem Vergleich

schlecht ein Vergleich ist. Er zeigt lediglich eine Tendenz und ist auch von den Spektren die verglichen werden sollen abhängig.

Tabelle 6-3: Ergebnisse des Spektren-Vergleichs am Beispiel von SAIB

Es sind die Ergebnisse einiger Vergleiche zwischen gemessenen und berechneten Spektren aufgelistet. Die Abkürzungen NF und F stehen für nicht fragmentiertes und fragmentiertes Spektrum, MN_0,001 für ein multinomial simuliertes Spektrum mit Schwellwert 0,001, FT_measRes für ein 2D-FT simuliertes Spektrum mit gemessener Auflösung und $d(Q, R)$ für die Gesamt-Distanz. Die Werte $t_{Distance}$ und t_{ppm} stellen die eingestellten Schwellwerte dar.

$t_{distance}$	t_{ppm}	Spektrum 1	Spektrum 2	$d(Q, R)$	M	MM	D/I
0,5	/	NF	F	0,04720	370	2	1364/21
1	0,5			0,20558	270	0	1466/121
0,5	/	F	Scan40	0,05313	263	0	1473/26
1	0,5			0,28213	222	0	1514/67
0,5	/	NF	MN_0.001	0,07784	108	119	117/283
1	0,5			0,23609	86	0	258/305
0,5	/	NF	FT_measRes	0,06959	93	0	151/298
1	0,5			0,14440	60	0	184/331

6.3.2 Analyse der Jeffrey-Matusita Distanz

Bei der Jeffrey-Matusita Distanz haben die zwei Parameter m/z und Auflösung einen Einfluss darauf ob ein Peak-Paar als Match klassifiziert wird oder nicht. In Abbildung 6-10 ist der Verlauf der JM-Distanz für verschiedene ΔR und $\Delta m/z$ – Werte in Abhängigkeit von Auflösung und nomineller Masse dargestellt. Dabei gibt $\Delta R = 0, 50.000, 100.000$ und $\Delta m/z = 0, 10^{-3}, 10^{-4}, 10^{-5}$ die Differenz in Auflösung und Masse von jeweils zwei zu vergleichenden Peaks wieder.

Aus der Abbildung geht hervor, dass die beiden Parameter einen sehr unterschiedlichen Einfluss auf die JM-Distanz besitzen. Eine größere Differenz in der Auflösung bei konstantem $\Delta m/z$ führt unabhängig vom m/z -Wert zu einem geringen Anstieg der JM-Distanz bei sinkender Auflösung. Eine zunehmende Massendifferenz hingegen führt zu einem starken Anstieg der JM-Distanz bei sinkender Masse. Für einen Vergleich von zwei Peaks bedeutet das: Ist die Masse beider Peaks sehr klein, die Massendifferenz aber relativ hoch, steigt die JM-Distanz unabhängig von der Auflösung. Ebenso führt eine geringe Auflösung unabhängig von der Masse zu einem Anstieg der Distanz.

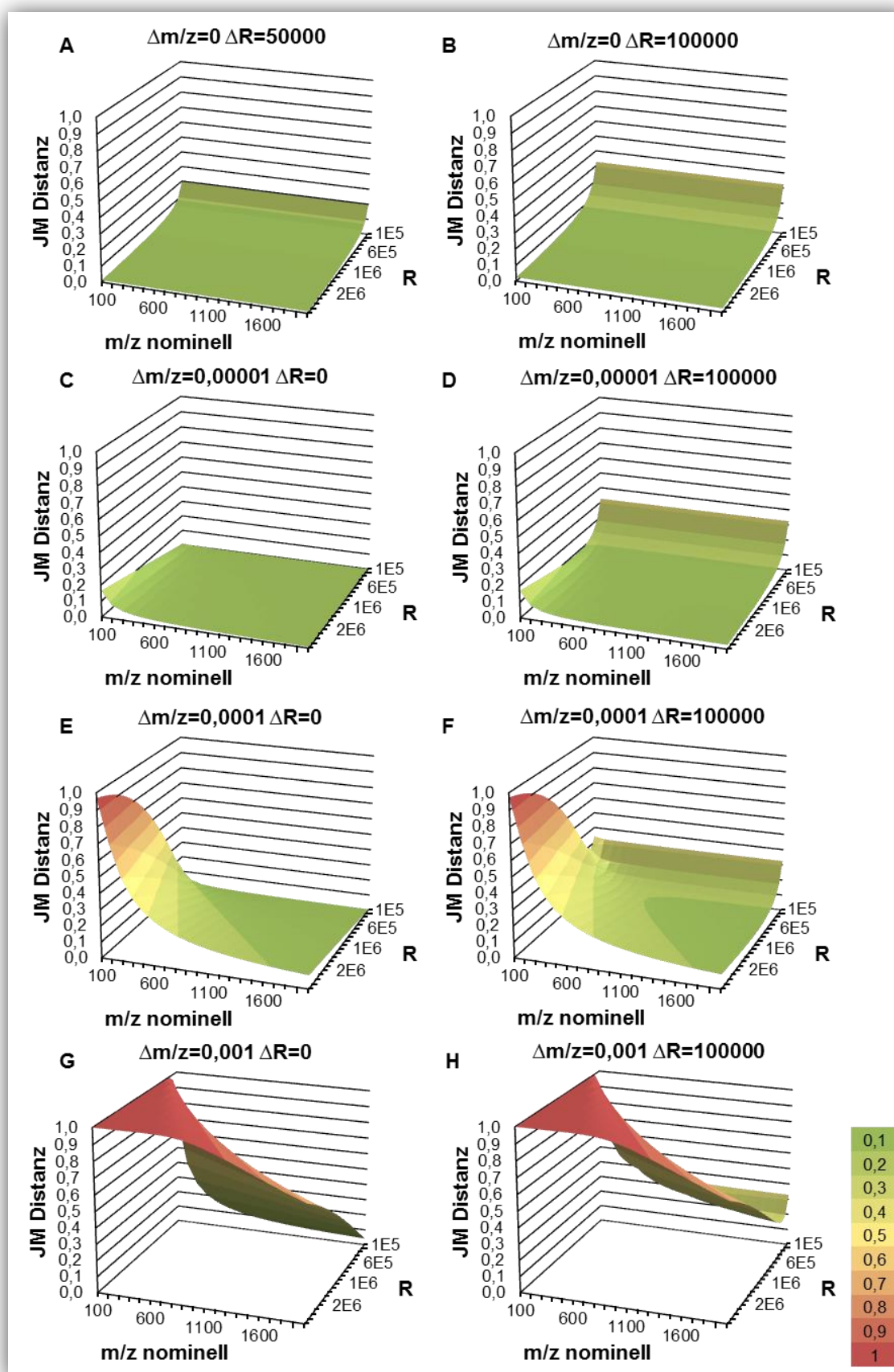


Abbildung 6-10: Jeffrey Matusita Distanz Analyse

(Beschreibung siehe Text)

7 Auswertung und Diskussion

7.1 Simulation der Spektren

7.1.1 Vergleich der Eingabe- und Ausgabe-Parameter

Für die Simulation von Massenspektren wurden die multinomiale (siehe Kapitel 3.4.2) und die 2D-FT Methode (siehe Kapitel 3.4.3) vorgestellt. In Tabelle 2-1 sind die benötigten Eingabeparameter und in Tabelle 7-2 die Ausgabewerte beider Methoden dargestellt. Beide Methoden benötigen mindestens die Summenformel eines Moleküls und deren Intensität. Die Intensität ist nicht notwendigerweise für das Berechnen der Isotopenmuster notwendig, wird aber um die gemessenen Spektren auch in Originalintensität darzustellen mit angegeben.

Tabelle 7-1: Vergleich der Eingabeparameter der multinomialen und 2D-FT Methode

Für die beiden Methoden werden unterschiedliche Parameter benötigt. Die Symbole zeigen die Notwendigkeit des jeweiligen Parameters an: ✓ - notwendig, × - nicht notwendig, (✓) – nicht für die reine Berechnung der Isotopenmuster notwendig..

	multinomial	2D-FT
Summenformel	✓	✓
Intensität	(✓)	(✓)
Auflösung	×	✓
Ladung	✓	✓
Schwellwert	✓	×

Beide Methoden liefern eine Summenformel, die Intensitäten und m/z -Werte der berechneten Peaks. Da die multinomiale Methode auf der Berechnung mit Hilfe von Isotopenwahrscheinlichkeiten basiert, können hier zusätzlich Informationen über die in der Summenformel enthaltenen Isotope angegeben werden. Bei der 2D-FT Methode ist dies nicht möglich, jedoch wird hier die Berechnung auf Grundlage der Auflösung durchgeführt, sodass zum einen das berechnete Spektrum ein Profilspektrum darstellt und zum anderen die Auflösung erhalten bleibt.

Tabelle 7-2: Vergleich Ausgabeparameter der multinomialen und 2D-FT- Methode

Die beiden Methoden haben unterschiedliche Ausgabewerte. Die Symbole zeigen an, ob die jeweiligen

Werte berechnet werden: ✓ - ja, ✗ - nein.

	multinomial	2D-FT
Summenformel	✓	✓
Isotopenformel	✓	✗
Intensität	✓	✓
m/z	✓	✓
Auflösung	✗	✓

Ein weiterer Unterschied der beiden Methoden ist, dass die multinomiale Methode ein Linienspektrum und die 2D-FT Methode ein Profilspektrum erzeugt. Die einzelnen Peaks des Profilspektrums haben die Form einer Gauß-Funktion. Aufgrund dieser Tatsache wurde es, um den Speicherbedarf zu minimieren, in ein Linienspektrum konvertiert. Das ursprüngliche Ergebnis kann jedoch, durch erneutes berechnen der Gauß-Kurve über jeden einzelnen Peak wieder hergestellt werden.

7.1.2 Laufzeit, Effizienz und Genauigkeit

Bei den Laufzeitmessungen stellte sich heraus, dass der multinomiale Algorithmus mit steigender Anzahl an Atomen und damit zunehmender Masse exponentiell bzw. wenn Elemente mit sechs oder mehr Isotopen in der Summenformel auftreten, faktoriell wächst. So benötigt der Algorithmus im *worst-case* Fall

$$O\left(\binom{n_{e_1} + i_{e_1} - 1}{i_{e_1} - 1} \cdot \dots \cdot \binom{n_{e_k} + i_{e_k} - 1}{i_{e_k} - 1}\right) \quad (7.1)$$

wobei n_{e_k} die Anzahl der Atome und i_{e_k} die Anzahl der Isotope des Elements e_k in der Summenformel darstellt. Eine Ausnahme bildet die Berechnung der Isotopenverteilungen von Molekülen mit ausschließlich mono- oder diisotopischen Elementen, bei denen die Laufzeit annähernd konstant bleibt (vgl. Abbildung 6-4). Aus diesem Grund ist der multinomiale Algorithmus mit seiner aktuellen Implementierung nicht für die Berechnung von Isotopenmustern von Summenformeln, welche polyisotopische Elemente mit mehr als 140 Atomen enthalten, geeignet. Des Weiteren ist die Rechenzeit weniger stark vom gesetzten Schwellwert abhängig, da bevor dieser im Algorithmus berücksichtigt wird, erst einmal alle möglichen Isotopenverteilungen mit Hilfe von Algorithmus 5.1 und 5.2 berechnet werden müssen. Dieser Teil des

Algorithmus nimmt die meiste Zeit in Anspruch. Bei kleineren Summenformeln, wie zum Beispiel bei den Spektren von SAIB kann die multinomiale Methode jedoch auch schneller als die 2D-FT Methode sein (vgl. Tabelle 6-1 und Tabelle 6-2).

Vorteil der multinomialen Methode ist, dass unabhängig von der Auflösung³⁸, alle Isotopenpeaks einer Summenformel oberhalb des gesetzten Schwellwertes berechnet werden. Es werden keine aggregierten Peaks gebildet und für jeden Peak erhält man Informationen über die genaue Zusammensetzung der Summenformel, d.h. welche Isotope enthalten sind. Die 2D-FT Methode ist besonders gut für die Berechnung von Isotopenmustern von Molekülen mit vielen polyisotopischen Elementen, da die Laufzeit auch bei einer großen Anzahl von Atomen annähernd konstant bleibt (vgl. Abbildung 6-8). Des Weiteren werden, abhängig von der Auflösung, Massendifferenz und Häufigkeit, Peaks als aggregierte Peaks zusammengefasst. In Abbildung 7-1 ist beispielhaft ein Peak bei m/z 865,4256 dargestellt. In grün ist der mit der multinomialen Methode (Schwellwert 0,001) und in blau das mit der 2D-FT Methode (gemessene Auflösung) berechnete Peak dargestellt. Es wurden zusätzlich zum Profilspektrum die einzelnen Peaks dargestellt. Es ist erkennbar, dass mit der multinomialen Methode vier einzelne Isotopenpeaks berechnet wurden und die 2D-FT Methode direkt den aggregierten Peak ausgibt. Wird eine höhere Auflösung gewählt, können jedoch auch bei der 2D-FT Methode die einzelnen Isotopenpeaks berechnet werden.

³⁸ Bei der Berechnung des Linienspektrums mit der multinomialen Methode kann man auch von einer Berechnung bei unendlich großer Auflösung sprechen.

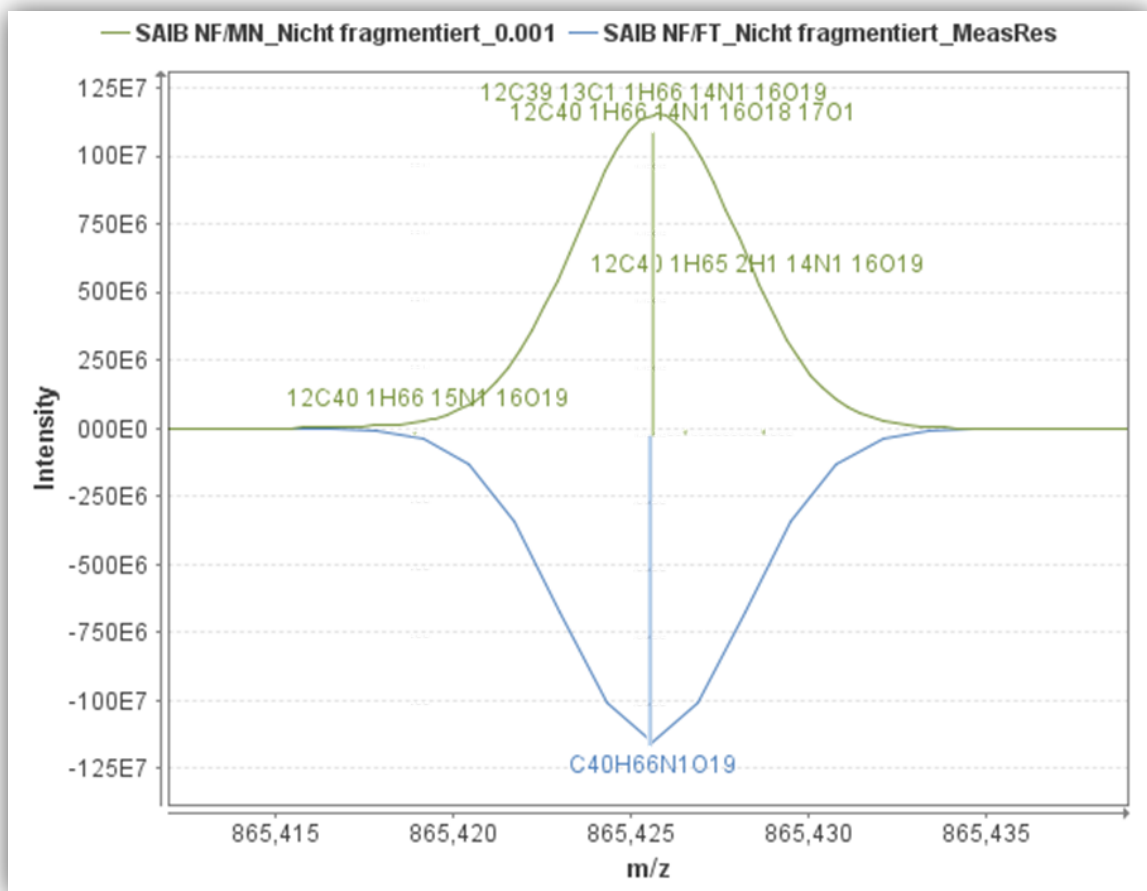


Abbildung 7-1: Vergleich eines mit der multinomialen und 2D-FT Methode berechneten Peaks

Ein Peak aus dem berechneten Spektrum des nicht fragmentierten Spektrums von SAIB. In grün ist das Ergebnis der multinomialen Methode (Schwellwert 0.001) und in blau das der 2D-FT Methode (gemessene Auflösung) dargestellt. Während die multinomiale Methode jeden Isotopenpeak einzeln berechnet gibt die 2D-FT Methode direkt den aggregierten Peak aus.

Vergleicht man das komplette multinomial berechnete Spektrum mit dem nach der 2D-FT Methode berechnete Spektrum von SAIB in Abbildung 7.1 erkennt man eine große Ähnlichkeit zwischen den beiden Ergebnissen. Nicht nur die Position der Peaks und die Isotopenmuster stimmen gut überein sondern auch die Intensitäten der M+1, M+2,... Peaks. Die Intensitäten der jeweiligen Molekülionen-Peaks sind gleich der Intensität des entsprechenden Peaks im gemessenen Spektrum. Die berechneten m/z -Werte stimmen bei beiden Methoden exakt überein, da die gleiche Isotopentabelle (siehe Anhang Tabelle A-1) verwendet wurde. Sie sind ebenfalls mit der berechneten Masse in der CSV Datei identisch.

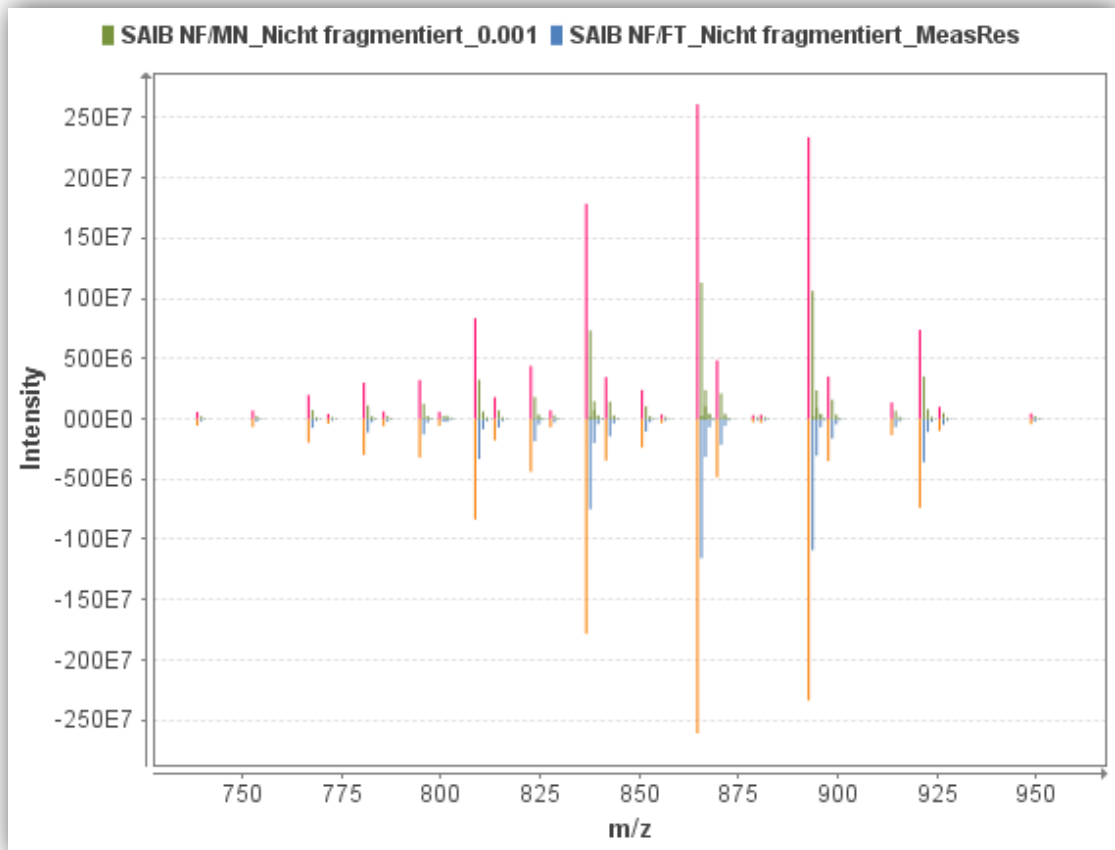


Abbildung 7-2: Vergleich 2D-FT Spektrum mit multinomial simuliertem Spektrum von SAIB

Das nicht fragmentierte Spektrum von SAIB wurde mit der multinomialen Methode (Schwellwert 0,001) und mit der 2D-FT Methode (gemessene Auflösung) simuliert. In grün ist das multinomiale und in blau das 2D-FT Spektrum dargestellt. In Magenta bzw. Orange sind jeweils die Molekül-Ionen-Peaks hervorgehoben.

Der größte Unterschied zwischen den beiden Methode, abgesehen von der Laufzeit, besteht in der Anzahl berechneter Peaks pro Summenformel (vgl. Abbildung 6-2 bzw. Abbildung 6-6). Abhängig von Schwellwert bzw. Auflösung werden unterschiedlich viele Isotopenpeaks berechnet. Da ein Massenspektrometer jedoch technisch bedingt ein gewisses Rauschen aufweist, können Peaks mit sehr niedriger Intensität nicht mehr detektiert werden bzw. gehen im Rauschen unter. Aufgrund dieser Tatsache müssen auch nicht alle Isotopenpeaks simuliert werden um sie mit einem gemessenen Spektrum vergleichen zu können. In Abbildung 7-3 ist dies an dem Basispeak des nicht fragmentierten Spektrum dargestellt. Es sind nur die ersten fünf Isotopen-Peaks des berechneten Spektrums (2D-FT Methode, gemessene Auflösung) im gemessenen Spektrum vorhanden. Die trifft für die meisten Peaks zu. Daher würde die Berechnung von fünf Peaks pro Summenformel für die vorliegenden Spektren ausreichen. Die Anzahl der gemessenen Isotopenpeaks ist jedoch von der elementaren Zusammensetzung sowie der Anzahl der Ionen im Massenspektrum abhängig.

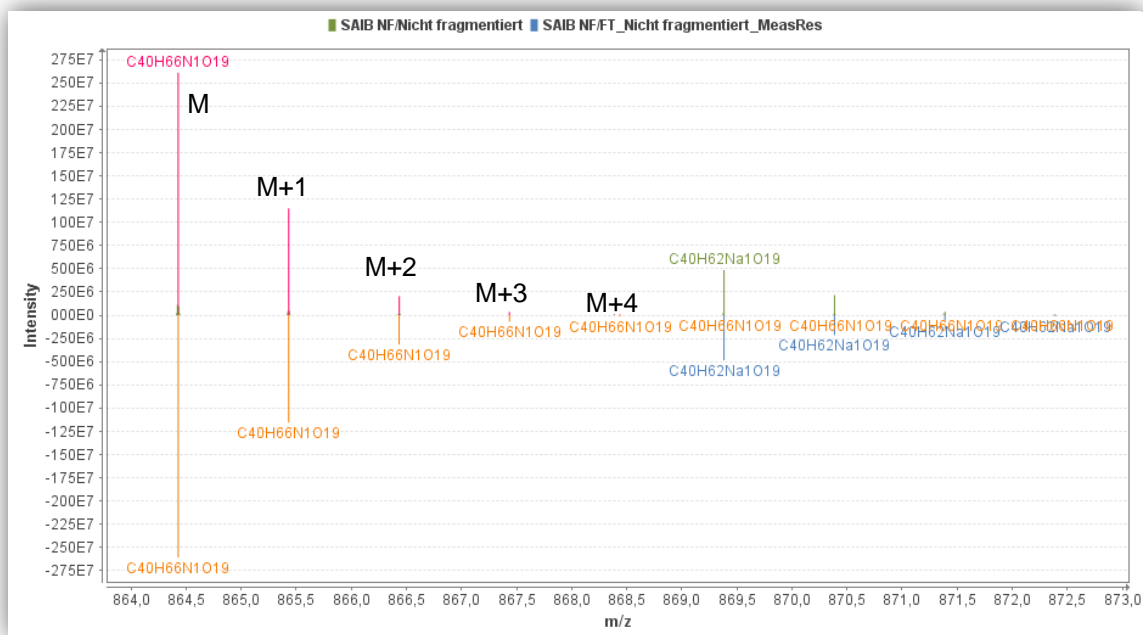


Abbildung 7-3: Vergleich eines berechneten und gemessenen Isotopenmusters

Es ist das Isotopenmuster des Basispeaks des nicht fragmentierten Spektrums (in Magenta hervorgehoben) und das zugehörige mit der 2D-Methode berechnete Isotopenmuster (in Orange hervorgehoben) dargestellt. Die ersten fünf Peaks (M bis M+4) des berechneten Isotopenmusters sind auch im gemessenen Spektrum vorhanden.

7.1.3 Parameterwahl

Mit Hilfe der Laufzeitmessungen sowie der Beispiel-Simulationen und der Analyse der berechneten Spektren können zusammenfassend allgemeine Aussagen über die Parameterwahl für die multinomiale und 2D-FT Methode getroffen werden. In Tabelle 7-3 sind noch einmal die wichtigsten Fakten zur Parameter und Methodenwahl dargestellt. Als Standard-Schwellwert für die multinomiale Methode wird 0,001 gewählt. Mit diesem Wert werden ausreichend Isotopenpeaks für einen Vergleich mit einem gemessenen Spektrum berechnet. Ein kleinerer Wert berechnet mehr, ein größerer Wert weniger Peaks (vgl. Tabelle 6-1). Für die 2D-FT Methode wird, wenn vorhanden vorrangig die gemessene Auflösung empfohlen. Ist diese nicht vorhanden kann eine der approximierten Auflösungsfunktionen gewählt werden. Im allgemeinen gilt, je kleiner die Auflösung desto weniger Peaks, werden berechnet und desto mehr Isotopenpeaks innerhalb einer nominellen Masse werden als aggregierter Peaks ausgegeben. In Abbildung 7-4 ist das Erscheinungsbild eines Peaks bei verschiedenen Auflösungen dargestellt.

Tabelle 7-3: Zusammenfassung und Parameterwahl für die Simulation von Spektren

	multinomial	2D-FT
Molekülgröße	bis 130 Atome für Polyisotopische Elemente wächst exponentiell bis faktoriell	egal
Laufzeit	Ausnahme: mono- und diisotopische Moleküle	annähernd konstant
Form	Linienpektrum einzelne Isotopenpeaks	Profilspektrum aggregierte Isotopenpeaks
Ladung	je nach Ionisationsmethode	je nach Ionisationsmethode
Schwellwert	je kleiner, desto mehr berechnete Peaks. Standard 0.001	/
Auflösung	/	Je höher desto mehr berechnete Peaks. Standard: gemessene Auflösung (wenn vorhanden), sonst approximierte

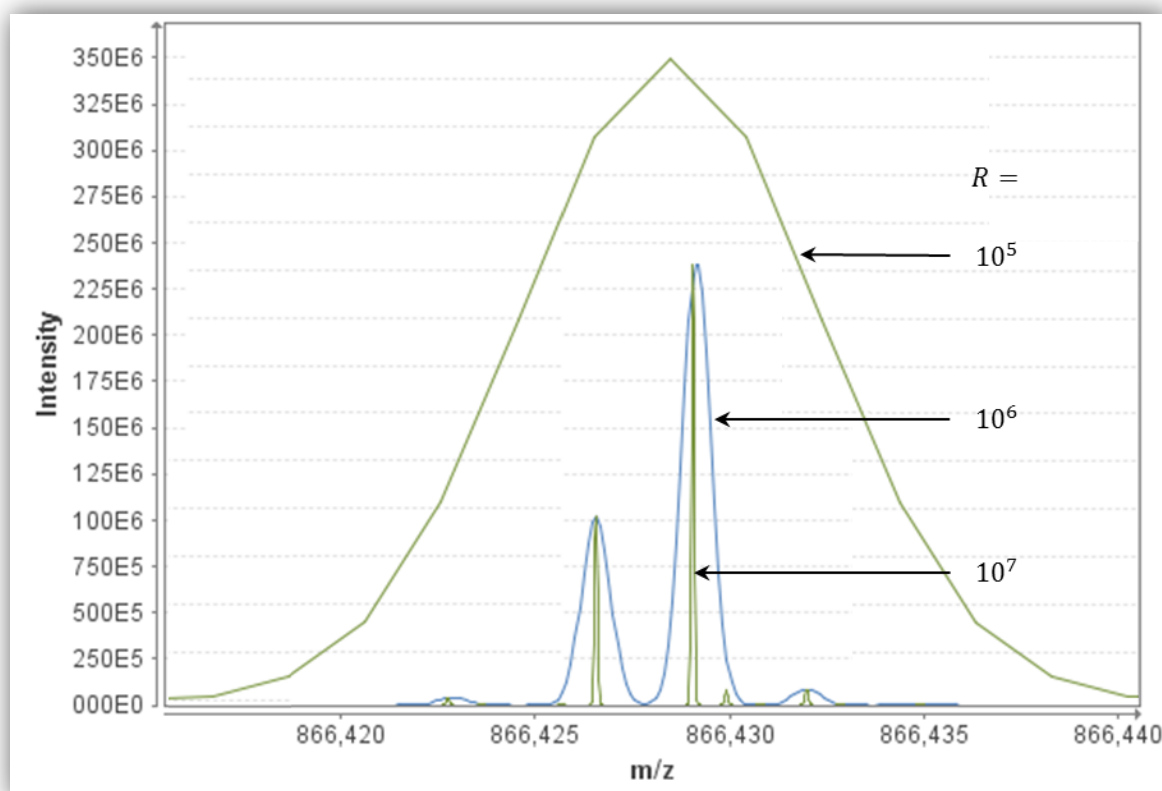


Abbildung 7-4: Ein mit 2D-FT simulierter Isotopenpeak mit verschiedenen Auflösungen

Es wurden die Auflösungen 10^5 , 10^6 und 10^7 verwendet. Es ist gut die verschiedenen Aggregationsstufen der einzelnen Isotopenpeaks zu erkennen. Während bei 10^5 nur ein einziger Peak berechnet wird sind es bei 10^7 sieben verschiedene.

7.2 Vergleich von Spektren

Der Algorithmus für den Vergleich von Spektren wurde in Kapitel 5.3 vorgestellt. Er basiert auf dem paarweisen Vergleich von Peaks. Die Anzahl an Matches gibt direkten Aufschluss darüber, welche Peaks die beiden Spektren gemeinsam haben. Insertionen und Deletionen sind Peaks die zu keinem Peak zugeordnet werden konnten. Eine besondere Rolle nehmen die Mismatches ein. Sie können auftreten, wenn zum Beispiel durch unterschiedliche Auflösungen zweier Peaks in dem einen Spektrum nur ein Peak und in dem anderen mehrere Peaks gemessen wurden.

Die Verwendung der JM-Distanz ist dann von Vorteil wenn die zu vergleichenden Spektren ähnliche Auflösungen besitzen. Die besten Ergebnisse wurden bei der Verwendung der gemessenen Auflösung erzielt (vgl. Tabelle 6-3). Des Weiteren ist die JM-Distanz für Peaks mit sehr großem m/z -Wert weniger geeignet. Aus Abbildung 6-10 geht hervor, dass sie bei zunehmender Masse immer mehr abnimmt und in ein asymptotisches Verhalten übergeht. Das führt zu immer kleiner werdenden Unterschieden der JM-Distanz bei zunehmender Masse und gleichbleibender Auflösungs- und Massendifferenz.

Ist die JM-Distanz nicht anwendbar, kann auf die ppm-Distanz ausgewichen werden. Die Peaks werden dabei nur anhand ihres ppm-Wertes verglichen. Dabei gilt, dass mit zunehmender Masse größere Massendifferenzen toleriert werden. Dies ist wichtig, da bei einem Massenspektrometer die Genauigkeit ebenfalls mit zunehmender Masse sinkt.

7.2.1 Parameterwahl

Welche Werte man für den Schwellwert $t_{distance}$ und t_{ppm} wählt hängt davon ab, wie viel Abweichung man zwischen zwei Peaks zulassen möchte um sie gerade noch als Match einzustufen. Für den Schwellwert $t_{distance}$ für die Verwendung der JM-Distanz haben sich Werte zwischen 0,3 und 0,5 als praktikabel erwiesen. Höhere Werte erhöhen die Anzahl gefundener Mismatches. Die in Abbildung 6-10 abgebildeten Plots können verwendet werden, um anwendungsspezifische Schwellwerte abzuleiten.

Für den Standardwert Standard ppm-Schwellwert wurde 0,5 gewählt. Abhängig von den zu vergleichenden Spektren kann ein höherer oder niedrigerer Schwellwert sinnvoll sein. Werden zum Beispiel zwei simulierte Spektren verglichen sollte ein sehr kleiner Wert gewählt werden, da die Massen einzelner Isotopenpeaks nicht oder nur sehr wenig voneinander abweichen.

8 Zusammenfassung und Ausblick

In dieser Arbeit wurden zwei Algorithmen zur Simulation von Massenspektren auf der Basis von Summenformeln vorgestellt: Die multinomiale Methode und die 2D-FT Methode. Des Weiteren wurde eine Möglichkeit vorgestellt Spektren, egal ob simuliert oder experimentell bestimmt, miteinander zu vergleichen. Die Algorithmen wurden anschließend in JAVA implementiert und die grafische Oberfläche SEAT dazu entwickelt.

Zusätzlich zu dem vorliegenden Dokument ist eine CD beigelegt, auf der das Programm SEAT und die verwendeten Massenspektren von SAIB enthalten sind. Ebenfalls ist ein Handbuch als Teil dieser Arbeit beigelegt. Das Programm ist voll funktionstüchtig und bietet neben der reinen Berechnung der Spektren und dem Vergleich von Spektren weitere hilfreiche Funktionen zur grafischen Darstellung sowie Import und Export von Dateien.

Als eine Verbesserung des Programms ist die effizientere Berechnung der simulierten Spektren mit Hilfe des multinomialen Algorithmus zu nennen. Des Weiteren ist eine Erweiterung des Vergleichs-Algorithmus sinnvoll, sodass zum Beispiel automatisch mit Hilfe der berechneten Spektren experimentell bestimmte Spektren von Rausch-Peaks oder Seitenpeaks bereinigt werden. Eine weitere Möglichkeit ist die Erweiterung der Analyse der verschiedenen Scans aus einem Kollisionsexperiment. und die verwendeten Massenspektren von SAIB enthalten sind.

Literaturverzeichnis

Primär- und Sekundärliteratur

- [1] Thomson, Joseph John: *Cathode Rays*. In: Philosophical Magazine 5(1897), 44, S. 293-316. - ISSN 1478-6443
- [2] Aston, Francis W.: *The constitution of atmospheric neon*. In: Philosophical Magazine 6(1920), 39, S. 449-455. - ISSN 0031-8086
- [3] Marshall, Alan G.; Rodgers, Ryan P.: *Petroleomics: Chemistry of the underworld*. In: PNAS 105(2008), 47, S. 18090–18095. - ISSN 1091-6490
- [5] Gross, Jürgen H.: *Massenspektrometrie: Ein Lehrbuch*. - 1. Aufl. - Berlin Heidelberg: Springer Verlag, 2011, S. 7
- [7] Yates III, John R.: *A century of mass spectrometry: from atoms to proteomes*. In: Nature Methods 8(2011), 8, S. 633 - 637. - ISSN 1548-7091
- [8] Watson, Throck J.; Sparkman, David O.: *Introduction to Mass Spectrometry: Instrumentation, Applications and Strategies for Data Interpretation*. - 4. Aufl. - Chichester: John Wiley & Sons Ltd., 2008
- [9] de Hoffmann, Edmond; Stroobant, Vincent: *Mass Spectrometry: Principles and Applications*. - 3. Aufl. - Chichester: John Wiley & Sons Ltd., 2007, S. 2f
- [10] Murray, Kermit K.; Boyd, Robert K.; et al.: *Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013)*. In: Pre Appl. Chem. 85(2013), 7, S. 1515-1609. - ISSN 1365-3075
- [12] Dole, Malcolm; Mack, L. L.; Hines, R. L.: *Molecular Beams of Macroions*. In: The Journal of Chemical Physics 49(1968), 1, S. 2240-2249. - ISSN 0021-9606
- [13] Rohner, Tatiana C.; Lion, Niels; Girault, Hubert H.: *Electrochemical and theoretical aspects of electrospray ionisation*. In: Physical Chemistry Chemistry Physics 6(2004), 5, S. 3056-3068. - ISSN 1463-9076
- [14] Fenn, John B.; Mann, Matthias; et al.: *Electrospray Ionization for Mass Spectrometry of Large Biomolecules*. In: Science 246(1989), 4926, S. 64-71. - ISSN 0036-8075
- [15] Gomez, Alessandro; Tang, Keqi: *Charge and fission of droplets in electrostatic sprays*. In: Physics of Fluids 6(1994), 1, S. 404-414. - ISSN 1089-7666

- [16] Desiderio, Dominic M.; Nibbering, Nico M.: *Mass spectrometry: Instrumentation, Interpretation and Applications*. - 1. Aufl. - Hoboken: John Wiley & Sons, Inc., 2009
- [17] Iribarne, J. V.; Thomson, B. A.: *On the evaporation of small ions from charged droplets*. In: *The Journal of Chemical Physics* 64(1976), 6, S. 2287-2294. - ISSN 0021-9606
- [18] Comisarov, Melvin B.; Marshall, Alan G.: *Fourier transform ion cyclotron resonance spectroscopy*. In: *Chemical Physics Letter* 25(1974), 2, S. 282-283. - ISSN 0009-2614
- [19] Marshall, Alan G.; Hendrickson, Christopher L.; Jackson, George S.: *Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer*. In: *Mass Spectrometry Reviews* 17(1998), 1, S. 1-35. - ISSN 1098-2787
- [20] Scigelove, Michaela; Hornshaw, Martin; et al.: *Fourier Transform Mass Spectrometry*. In: *Molecular and Cellular Proteomics* 10(2011), 7, S. 1-19. - ISSN 1535-9484
- [21] Guan, Shenheng; Marshall, Alan G.: *Stored waveform inverse Fourier transform (SWIFT) ion excitation in trapped-ion mass spectrometry: theory and applications*. In: *International Journal of Mass Spectrometry and Ion Processes* 157-158(1996), 1, S. 5-37. - ISSN 1387-3806
- [25] KICKELBICK, Guido: *Chemie für Ingenieure*. - 1. Aufl. - München: Pearson Studium, 2008, S. 40
- [26] Berglund, Michael; Wieser, Michael E.: *Isotopic compositions of the elements 2009 (IUPAC Technical Report)*. In: *Pure Appl. Chem.* 83(2011), 2, S. 397-410. - ISSN 1365-3075
- [30] Coplen, Tyler B.: *Atomic Weights of the Elements 1999 (IUPAC Technical Report)*. In: *Pure Appl. Chem.* 73(2001), 4, S. 667-683. - ISSN 1365-3075
- [32] Riedel, Erwin: *Allgemeine und anorganische Chemie*. - 10. Aufl. - Berlin: Walter de Gruyter GmbH & Co. KG, 2010, S. 9 ff.
- [33] Busch, Kenneth L.: *A Glossary for Mass Spectrometry*. In: *Mass Spectrometry* 16(2001), 11, S. 28-31. - ISSN 0887-6703
- [34] Valkenburg, Dirk; Mertens, Inge; et al.: *The Isotopic Distribution Conundrum*. In: *Mass Spectrometry Reviews* 12(2012), , S. 96-109. - ISSN 1098-2787

- [35] Yergey, James A.: *A General Approach to Calculating Isotopic Distributions For Mass Spectrometry*. In: *International Journal of Mass Spectrometry* 52(1983), 2/3, S. 337-349. - ISSN 1387-3806
- [36] Rockwood, Alan L.: *Relationship of Fourier Transforms to Isotope Distribution Calculations*. In: *Rapid Communications in Mass Spectrometry* 9(1995), 1, S. 103-105. - ISSN 1097-0231
- [37] Rockwood, Alan L.; Van Orden, Steven L.; Smith, Richard D.: *Rapid Calculation of Isotope Distributions*. In: *Analytical Chemistry* 67(1995), 15, S. 2699-2704. - ISSN 1520-6882
- [38] Rockwood, Alan L.; Van Orden, Steven L.; Smith, Richard D.: *Ultrahigh Resolution Isotope Distribution Calculations*. In: *Rapid Communications in Mass Spectrometry* 10(1996), 1, S. 54-59. - ISSN 1097-0231
- [39] Fernandez-de-Cossio, Jorge: *Efficient Packing Fourier-Transform Approach for Ultrahigh Resolution Isotopic Distribution Calculations*. In: *Analytical Chemistry* 82(2010), 5, S. 1759-1765. - ISSN 1520-6882
- [40] Fernandez-de-Cossio, Jorgo: *Computation of the Isotopic Distribution in Two Dimensions*. In: *Analytical Chemistry* 82(2010), 15, S. 6726-6729. - ISSN 0003-2700
- [42] Reynolds, C. R.; Chappel, I. C.: *Sucrose Acetate Isobutyrate (SAIB): Historical Aspects of its use in Beverages and a Review of Toxicity Studies Prior to 1988*. In: *Food and Chemical Toxicology* 36(1998), 2, S. 81-93. - ISSN 0278-6915
- [43] Kelleher, Jerome; O'Sullivan, Barry: *Generating All Partitions: A Comparison Of Two Encodings*. In: *Cornell University Library* (2009), , S. 1-40. - ISSN
- [45] Hansen, Michael Edberg; Smedsgaard, Jorn: *A new Matching Algorithm for High Resolution Mass Spectra*. In: *J. Am. Soc. Mass Spectrom.* 15(2004), 8, S. 1173 - 1180. - ISSN 1879-1123
- [46] Kailath, Thomas: *The Divergence and Bhattacharyya Distance Measures in Signal Selection*. In: *IEEE Transactions on Communication Technology* 15(1967), 1, S. 52-60. - ISSN 0090-6778
- [47] Fukunaga, Keinosuke: *Introduction to Statistical Pattern Recognition*. - 2. Aufl. - San Diego: Academic Press, 1990, S. 188f
- [48] Bruzzone, Lorenzo; Roli, Fabio; Serpico, Sebastiano B.: *An Extension in the*

Jeffreys-Matusita Distance to Multiclass Cases for Feature Selection. In: IEEE Transactions on Geoscience and Remote Sensing 33(1995), 6, S. 1318-1321. - ISSN 0196-2892

- [51] Rockwood, Alan L.: *Deconvoluting Isotopic Distributions to Evaluate Parent/Fragment Ion Relationships*. In: Rapid Communications in Mass Spectrometry 11(1997), 3, S. 241-248. - ISSN 1097-0231

Onlinequellen

- [4] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: mass spectroscopy.
URL: < <http://goldbook.iupac.org/M03748.html> >, verfügbar am 19.. August, 2013
- [6] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: mass-to-charge ratio.
URL: < <http://goldbook.iupac.org/M03752.html> >, verfügbar am 21. Dezember, 2013
- [11] The Nobel Foundation, : Nobelprize.
URL: < http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2002/fenn-lecture.html >, verfügbar am 28. November, 2013
- [22] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: resolution.
URL: < <http://goldbook.iupac.org/R05318.html> >, verfügbar am 9. Dezember, 2013
- [23] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: unified atomic mass unit.
URL: < <http://goldbook.iupac.org/U06554.html> >, verfügbar am 11. Januar, 2014
- [24] McIntyre, Doug: *Effect of Resolution and Mass Accuracy on Empirical Formula Confirmation and Identification of Unknowns*.
URL: < <http://www.chem.agilent.com/Library/technicaloverviews/Public/5989-1052EN%2014-May-2004.pdf> >, verfügbar am 19. Dezember, 2013
- [27] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: isotopologue.
URL: < <http://goldbook.iupac.org/I03351.html> >, verfügbar am 16. Dezember, 2013
- [28] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: elementary charge.

- URL: < <http://goldbook.iupac.org/I03330.html> >, verfügbar am 16. Dezember, 2013
- [29] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: *Isotopomer*.
URL: < <http://goldbook.iupac.org/I03352.html> >, verfügbar am 10. Dezember, 2013
- [31] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: *atomic units*.
URL: < <http://goldbook.iupac.org/A00504.html> >, verfügbar am 29.09. September, 2013
- [41]: *Zusatzstoffe - Was ist drin im Lebensmittel*.
URL: < <http://zusatzstoffe.e314.de/zusatzstoff/E444.html> >, verfügbar am 19. September, 2013
- [44] Fernandez-de-Cossio, Jorge: *Deuterium*.
URL: < <http://bioinformatica.cigb.edu.cu/isotopica/deuterium.html> >, verfügbar am 17. Juli, 2013
- [49] Gates, Paul J.: *University of Bristol: Ions & Ionisation Techniques*.
URL: < <http://www.chm.bris.ac.uk/ms/newversion/eici-ionisation.htm> >, verfügbar am 18. September, 2005
- [50] Nic, Miloslav; Jirat, Jiri; Kosata, Bedrich: *IUPAC: Compendium of Chemical Terminology (Gold Book)* - Stichwort: *elementary charge*.
URL: < <http://goldbook.iupac.org/E02032.html> >, verfügbar am 31. Dezember, 2013

Anhang

A	Massen und Häufigkeiten der Elemente und deren Isotop	XVIII
B	Peak-Anzahl simuliertes fragmentiertes Spektrum	XXV
C	Laufzeiten Multinomial und 2D-FT Methode	XXVII

A Massen und Häufigkeiten der Elemente und deren Isotope

Tabelle A-1: Isotopenzusammensetzung sowie Massen der Elemente

Symbol	Element-name	Z	N	Isotopen-Masse [u]	Häufigkeit [%]	relative Häufigkeit der Isotope	relative Atommasse [u]
H	Wasserstoff	1	1	1,007825	100	0,999885	1,00795
H	Wasserstoff	1	2	2,014101	0,0115	0,000115	1,00795
He	Helium	2	3	3,016029	0,000137	0,00000134	4,002602
He	Helium	2	4	4,002603	100	0,99999866	4,002602
Li	Lithium	3	6	6,015122	8,21	0,0759	6,941
Li	Lithium	3	7	7,016004	100	0,9241	6,941
Be	Beryllium	4	9	9,012182	100	1	9,012182
B	Bor	5	10	10,012937	24,8	0,199	10,812
B	Bor	5	11	11,009306	100	0,801	10,812
C	Kohlenstoff	6	12	12	100	0,9893	12,0108
C	Kohlenstoff	6	13	13,003355	1,08	0,0107	12,0108
N	Stickstoff	7	14	14,003074	100	0,99636	14,00675
N	Stickstoff	7	15	15,000109	0,369	0,00364	14,00675
O	Sauerstoff	8	16	15,994915	100	0,99757	15,9994
O	Sauerstoff	8	17	16,999132	0,038	0,00038	15,9994
O	Sauerstoff	8	18	17,999116	0,205	0,00205	15,9994
F	Fluor	9	19	18,998403	100	1	18,998403
Ne	Neon	10	20	19,992402	100	0,9048	20,1798
Ne	Neon	10	21	20,993847	0,3	0,0027	20,1798
Ne	Neon	10	22	21,991386	10,22	0,0925	20,1798
Na	Natrium	11	23	22,989769	100	1	22,989769
Mg	Magnesium	12	24	23,985042	100	0,7899	24,3051
Mg	Magnesium	12	25	24,985837	12,66	0,1	24,3051
Mg	Magnesium	12	26	25,982593	13,94	0,1101	24,3051
Al	Aluminium	13	27	26,981538	100	1	26,981538
Si	Silicium	14	28	27,976927	100	0,92223	28,0855
Si	Silicium	14	29	28,976495	5,0778	0,04685	28,0855
Si	Silicium	14	30	29,97377	3,3473	0,03092	28,0855
P	Phosphor	15	31	30,973762	100	1	30,973762
S	Schwefel	16	32	31,972071	100	0,9499	32,067
S	Schwefel	16	33	32,971459	0,8	0,0075	32,067
S	Schwefel	16	34	33,967867	4,52	0,0425	32,067
S	Schwefel	16	36	35,967081	0,02	0,0001	32,067
Cl	Chlor	17	35	34,968853	100	0,7576	35,4528
Cl	Chlor	17	37	36,965903	31,96	0,2424	35,4528
Ar	Argon	18	36	35,967546	0,3379	0,003336	39,948

Ar	Argon	18	38	37,962776	0,0635	0,000629	39,948
Ar	Argon	18	40	39,962383	100	0,996035	39,948
K	Kalium	19	39	38,963706	100	0,932581	39,0983
K	Kalium	19	40	39,963999	0,0125	0,000117	39,0983
K	Kalium	19	41	40,961826	7,2167	0,067302	39,0983
Ca	Calcium	20	40	39,962591	100	0,96941	40,078
Ca	Calcium	20	42	41,958618	0,667	0,00647	40,078
Ca	Calcium	20	43	42,958769	0,139	0,00135	40,078
Ca	Calcium	20	44	43,955481	2,152	0,02086	40,078
Ca	Calcium	20	46	45,953693	0,004	0,00004	40,078
Ca	Calcium	20	48	47,952534	0,193	0,00187	40,078
Sc	Scandium	21	45	44,95591	100	1	44,95591
Ti	Titan	22	46	45,952629	11,19	0,0825	47,867
Ti	Titan	22	47	46,951764	10,09	0,0744	47,867
Ti	Titan	22	48	47,947947	100	0,7372	47,867
Ti	Titan	22	49	48,947871	7,34	0,0541	47,867
Ti	Titan	22	50	49,944792	7,03	0,0518	47,867
V	Vanadium	23	50	49,947163	0,25	0,0025	50,9415
V	Vanadium	23	51	50,943964	100	0,9975	50,9415
Cr	Chrom	24	50	49,94605	5,187	0,04345	51,9962
Cr	Chrom	24	52	51,940512	100	0,83789	51,9962
Cr	Chrom	24	53	52,940654	11,339	0,09501	51,9962
Cr	Chrom	24	54	53,938885	2,823	0,02365	51,9962
Mn	Mangan	25	55	54,93805	100	1	54,93805
Fe	Eisen	26	54	53,939615	6,37	0,05845	55,845
Fe	Eisen	26	56	55,934942	100	0,91754	55,845
Fe	Eisen	26	57	56,935399	2,309	0,02119	55,845
Fe	Eisen	26	58	57,93328	0,307	0,00282	55,845
Co	Cobalt	27	59	58,9332	100	1	58,9332
Ni	Nickel	28	58	57,935348	100	0,68077	58,6934
Ni	Nickel	28	60	59,930791	38,5198	0,26223	58,6934
Ni	Nickel	28	61	60,93106	1,6744	0,011399	58,6934
Ni	Nickel	28	62	61,928349	5,3388	0,036346	58,6934
Ni	Nickel	28	64	63,92797	1,3596	0,009255	58,6934
Cu	Kupfer	29	63	62,929601	100	0,6915	63,546
Cu	Kupfer	29	65	64,927794	44,57	0,3085	63,546
Zn	Zink	30	64	63,929147	100	0,4917	65,39
Zn	Zink	30	66	65,926037	57,37	0,2773	65,39
Zn	Zink	30	67	66,927131	8,43	0,0404	65,39
Zn	Zink	30	68	67,924848	38,56	0,1845	65,39
Zn	Zink	30	70	69,925325	1,27	0,0061	65,39
Ga	Gallium	31	69	68,925581	100	0,60108	69,723
Ga	Gallium	31	71	70,924705	66,367	0,39892	69,723
Ge	Germanium	32	70	69,92425	56,44	0,2057	72,61
Ge	Germanium	32	72	71,922076	75,91	0,2745	72,61

Ge	Germanium	32	73	72,923459	21,31	0,0775	72,61
Ge	Germanium	32	74	73,921178	100	0,365	72,61
Ge	Germanium	32	76	75,921403	20,98	0,0773	72,61
As	Arsen	33	75	74,921596	100	1	74,921596
Se	Selen	34	74	73,922477	1,79	0,0089	78,96
Se	Selen	34	76	75,919214	18,89	0,0937	78,96
Se	Selen	34	77	76,919915	15,38	0,0763	78,96
Se	Selen	34	78	77,91731	47,91	0,2377	78,96
Se	Selen	34	80	79,916522	100	0,4961	78,96
Se	Selen	34	82	81,9167	17,6	0,0873	78,96
Br	Brom	35	79	78,918338	100	0,5069	79,904
Br	Brom	35	81	80,916291	97,28	0,4931	79,904
Kr	Krypton	36	78	77,920387	0,61	0,00355	83,8
Kr	Krypton	36	80	79,916378	4	0,02286	83,8
Kr	Krypton	36	82	81,913485	20,32	0,11593	83,8
Kr	Krypton	36	83	82,914136	20,16	0,115	83,8
Kr	Krypton	36	84	83,911507	100	0,56987	83,8
Kr	Krypton	36	86	85,91061	30,35	0,17279	83,8
Rb	Rubidium	37	85	84,911789	100	0,7217	85,4678
Rb	Rubidium	37	87	86,909183	38,56	0,2783	85,4678
Sr	Strontium	38	84	83,913425	0,68	0,0056	87,62
Sr	Strontium	38	86	85,909262	11,94	0,0986	87,62
Sr	Strontium	38	87	86,908879	8,48	0,07	87,62
Sr	Strontium	38	88	87,905614	100	0,8258	87,62
Y	Yttrium	39	89	88,905848	100	1	88,905848
Zr	Zirkonium	40	90	89,904704	100	0,5145	91,224
Zr	Zirkonium	40	91	90,905645	21,81	0,1122	91,224
Zr	Zirkonium	40	92	91,90504	33,33	0,1715	91,224
Zr	Zirkonium	40	94	93,906316	33,78	0,1738	91,224
Zr	Zirkonium	40	96	95,908276	5,44	0,028	91,224
Nb	Niob	41	93	92,906378	100	1	92,906378
Mo	Molybdän	42	92	91,90681	61,5	0,1453	95,94
Mo	Molybdän	42	94	93,905088	38,33	0,0915	95,94
Mo	Molybdän	42	95	94,905841	65,98	0,1584	95,94
Mo	Molybdän	42	96	95,904679	69,13	0,1667	95,94
Mo	Molybdän	42	97	96,906021	39,58	0,096	95,94
Mo	Molybdän	42	98	97,905408	100	0,2439	95,94
Mo	Molybdän	42	100	99,907478	39,91	0,0982	95,94
Ru	Ruthenium	44	96	95,907599	17,56	0,0554	101,07
Ru	Ruthenium	44	98	97,905288	5,93	0,0187	101,07
Ru	Ruthenium	44	99	98,905939	40,44	0,1276	101,07
Ru	Ruthenium	44	100	99,904229	39,94	0,126	101,07
Ru	Ruthenium	44	101	100,905582	54,07	0,1706	101,07
Ru	Ruthenium	44	102	101,90435	100	0,3155	101,07
Ru	Ruthenium	44	104	103,90543	59,02	0,1862	101,07

Rh	Rhodium	45	103	102,905504	100	1	102,905504
Pd	Palladium	46	102	101,905608	3,73	0,0102	106,42
Pd	Palladium	46	104	103,904036	40,76	0,1114	106,42
Pd	Palladium	46	105	104,905084	81,71	0,2233	106,42
Pd	Palladium	46	106	105,903484	100	0,2733	106,42
Pd	Palladium	46	108	107,903894	96,82	0,2646	106,42
Pd	Palladium	46	110	109,905151	42,88	0,1172	106,42
Ag	Silber	47	107	106,905094	100	0,51839	107,8682
Ag	Silber	47	109	108,904756	92,9	0,48161	107,8682
Cd	Cadmium	48	106	105,906459	4,35	0,0125	112,412
Cd	Cadmium	48	108	107,904184	3,1	0,0089	112,412
Cd	Cadmium	48	110	109,903006	43,47	0,1249	112,412
Cd	Cadmium	48	111	110,904182	44,55	0,128	112,412
Cd	Cadmium	48	112	111,902757	83,99	0,2413	112,412
Cd	Cadmium	48	113	112,904401	42,53	0,1222	112,412
Cd	Cadmium	48	114	113,903358	100	0,2873	112,412
Cd	Cadmium	48	116	115,904755	26,07	0,0749	112,412
In	Indium	49	113	112,904061	4,48	0,0429	114,818
In	Indium	49	115	114,903879	100	0,9571	114,818
Sn	Zinn	50	112	111,904822	2,98	0,0097	118,711
Sn	Zinn	50	114	113,902782	2,03	0,0066	118,711
Sn	Zinn	50	115	114,903346	1,04	0,0034	118,711
Sn	Zinn	50	116	115,901744	44,63	0,1454	118,711
Sn	Zinn	50	117	116,902954	23,57	0,0768	118,711
Sn	Zinn	50	118	117,901606	74,34	0,2422	118,711
Sn	Zinn	50	119	118,903309	26,37	0,0859	118,711
Sn	Zinn	50	120	119,902197	100	0,3258	118,711
Sn	Zinn	50	122	121,90344	14,21	0,0463	118,711
Sn	Zinn	50	124	123,905275	17,77	0,0579	118,711
Sb	Antimon	51	121	120,903818	100	0,5721	121,76
Sb	Antimon	51	123	122,904216	74,79	0,4279	121,76
Te	Tellur	52	120	119,904021	0,26	0,0009	127,6
Te	Tellur	52	122	121,903047	7,48	0,0255	127,6
Te	Tellur	52	123	122,904273	2,61	0,0089	127,6
Te	Tellur	52	124	123,902819	13,91	0,0474	127,6
Te	Tellur	52	125	124,904425	20,75	0,0707	127,6
Te	Tellur	52	126	125,903306	55,28	0,1884	127,6
Te	Tellur	52	128	127,904461	93,13	0,3174	127,6
Te	Tellur	52	130	129,906223	100	0,3408	127,6
I	Iod	53	127	126,904468	100	1	126,904468
Xe	Xenon	54	124	123,905896	0,33	0,000952	131,29
Xe	Xenon	54	126	125,90427	0,33	0,00089	131,29
Xe	Xenon	54	128	127,90353	7,14	0,019102	131,29
Xe	Xenon	54	129	128,904779	98,33	0,264006	131,29
Xe	Xenon	54	130	129,903508	15,17	0,04071	131,29

Xe	Xenon	54	131	130,905082	78,77	0,212324	131,29
Xe	Xenon	54	132	131,904154	100	0,269086	131,29
Xe	Xenon	54	134	133,905395	38,82	0,104357	131,29
Xe	Xenon	54	136	135,907221	32,99	0,088573	131,29
Cs	Caesium	55	133	132,905447	100	1	132,905447
Ba	Barium	56	130	129,906311	0,148	0,00106	137,328
Ba	Barium	56	132	131,905056	0,141	0,00101	137,328
Ba	Barium	56	134	133,904503	3,371	0,02417	137,328
Ba	Barium	56	135	134,905683	9,194	0,06592	137,328
Ba	Barium	56	136	135,90457	10,954	0,07854	137,328
Ba	Barium	56	137	136,905821	15,666	0,11232	137,328
Ba	Barium	56	138	137,905241	100	0,71698	137,328
La	Lanthan	57	138	137,907107	0,09	0,0008881	138,9055
La	Lanthan	57	139	138,906348	100	0,9991119	138,9055
Ce	Cer	58	136	135,907145	0,209	0,00185	140,116
Ce	Cer	58	138	137,905991	0,284	0,00251	140,116
Ce	Cer	58	140	139,905434	100	0,8845	140,116
Ce	Cer	58	142	141,90924	12,565	0,11114	140,116
Pr	Praseodym	59	141	140,907648	100	1	140,907648
Nd	Neodym	60	142	141,907719	100	0,27152	144,24
Nd	Neodym	60	143	142,90981	44,9	0,12174	144,24
Nd	Neodym	60	144	143,910083	87,5	0,23798	144,24
Nd	Neodym	60	145	144,912569	30,5	0,08293	144,24
Nd	Neodym	60	146	145,913112	63,2	0,17189	144,24
Nd	Neodym	60	148	147,916889	21	0,05756	144,24
Nd	Neodym	60	150	149,920887	20,6	0,05638	144,24
Sm	Samarium	62	144	143,911995	11,48	0,0307	150,36
Sm	Samarium	62	147	146,914893	56,04	0,1499	150,36
Sm	Samarium	62	148	147,914818	42,02	0,1124	150,36
Sm	Samarium	62	149	148,91718	51,66	0,1382	150,36
Sm	Samarium	62	150	149,917271	27,59	0,0738	150,36
Sm	Samarium	62	152	151,919728	100	0,2675	150,36
Sm	Samarium	62	154	153,922205	85,05	0,2275	150,36
Eu	Europium	63	151	150,919846	91,61	0,4781	151,964
Eu	Europium	63	153	152,921226	100	0,5219	151,964
Gd	Gadolinium	64	152	151,919788	0,81	0,002	157,25
Gd	Gadolinium	64	154	153,920862	8,78	0,0218	157,25
Gd	Gadolinium	64	155	154,922619	59,58	0,148	157,25
Gd	Gadolinium	64	156	155,92212	82,41	0,2047	157,25
Gd	Gadolinium	64	157	156,923957	63	0,1565	157,25
Gd	Gadolinium	64	158	157,924101	100	0,2484	157,25
Gd	Gadolinium	64	160	159,927051	88	0,2186	157,25
Tb	Terbium	65	159	158,925343	100	1	158,925343
Dy	Dysprosium	66	156	155,924279	0,21	0,00056	162,5
Dy	Dysprosium	66	158	157,924405	0,35	0,00095	162,5

Dy	Dysprosium	66	160	159,925194	8,3	0,02329	162,5
Dy	Dysprosium	66	161	160,92693	67,1	0,18889	162,5
Dy	Dysprosium	66	162	161,926795	90,53	0,25475	162,5
Dy	Dysprosium	66	163	162,928728	88,36	0,24896	162,5
Dy	Dysprosium	66	164	163,929171	100	0,00139	162,5
Ho	Holmium	67	165	164,930319	100	1	164,930319
Er	Erbium	68	162	161,928775	0,42	0,00139	167,26
Er	Erbium	68	164	163,929197	4,79	0,01601	167,26
Er	Erbium	68	166	165,93029	100	0,33503	167,26
Er	Erbium	68	167	166,932045	68,22	0,22869	167,26
Er	Erbium	68	168	167,932368	79,69	0,26978	167,26
Er	Erbium	68	170	169,93546	44,42	0,1491	167,26
Tm	Thulium	69	169	168,934211	100	1	168,934211
Yb	Ytterbium	70	168	167,933894	0,41	0,00123	173,04
Yb	Ytterbium	70	170	169,934759	9,55	0,02982	173,04
Yb	Ytterbium	70	171	170,936322	44,86	0,1409	173,04
Yb	Ytterbium	70	172	171,936378	68,58	0,2168	173,04
Yb	Ytterbium	70	173	172,938207	50,68	0,16103	173,04
Yb	Ytterbium	70	174	173,938858	100	0,32026	173,04
Yb	Ytterbium	70	176	175,942568	40,09	0,12996	173,04
Lu	Lutetium	71	175	174,940768	100	0,97401	174,967
Lu	Lutetium	71	176	175,942682	2,66	0,02599	174,967
Hf	Hafnium	72	174	173,94004	0,46	0,0016	178,49
Hf	Hafnium	72	176	175,941402	14,99	0,0526	178,49
Hf	Hafnium	72	177	176,94322	53,02	0,186	178,49
Hf	Hafnium	72	178	177,943698	77,77	0,2728	178,49
Hf	Hafnium	72	179	178,944815	38,83	0,1362	178,49
Hf	Hafnium	72	180	179,946549	100	0,3508	178,49
Ta	Tantal	73	180	179,947466	0,012	0,9998799	180,9479
Ta	Tantal	73	181	180,947996	100	0,9998799	180,9479
W	Wolfram	74	180	179,946707	0,4	0,0012	183,84
W	Wolfram	74	182	181,948206	86,49	0,265	183,84
W	Wolfram	74	183	182,950224	46,7	0,1431	183,84
W	Wolfram	74	184	183,950933	100	0,3064	183,84
W	Wolfram	74	186	185,954362	93,79	0,2843	183,84
Re	Rhenium	75	185	184,952956	59,74	0,374	186,207
Re	Rhenium	75	187	186,955751	100	0,626	186,207
Os	Osmium	76	184	183,952491	0,05	0,0002	190,23
Os	Osmium	76	186	185,953838	3,9	0,0159	190,23
Os	Osmium	76	187	186,955748	4,81	0,0196	190,23
Os	Osmium	76	188	187,955836	32,47	0,1324	190,23
Os	Osmium	76	189	188,958145	39,6	0,1615	190,23
Os	Osmium	76	190	189,958445	64,39	0,2626	190,23
Os	Osmium	76	192	191,961479	100	0,4078	190,23
Ir	Iridium	77	191	190,960591	59,49	0,373	192,217

Ir	Iridium	77	193	192,962924	100	0,627	192,217
Pt	Platin	78	190	189,959931	0,041	0,00012	195,078
Pt	Platin	78	192	191,961035	2,311	0,00782	195,078
Pt	Platin	78	194	193,962664	97,443	0,3286	195,078
Pt	Platin	78	195	194,964774	100	0,3378	195,078
Pt	Platin	78	196	195,964935	74,61	0,2521	195,078
Pt	Platin	78	198	197,967876	21,172	0,07356	195,078
Au	Gold	79	197	196,966552	100	1	196,966552
Hg	Quecksilber	80	196	195,965815	0,5	0,0015	200,59
Hg	Quecksilber	80	198	197,966752	33,39	0,0997	200,59
Hg	Quecksilber	80	199	198,968262	56,5	0,1687	200,59
Hg	Quecksilber	80	200	199,968309	77,36	0,231	200,59
Hg	Quecksilber	80	201	200,970285	44,14	0,1318	200,59
Hg	Quecksilber	80	202	201,970626	100	0,2986	200,59
Hg	Quecksilber	80	204	203,973476	23	0,0687	200,59
Tl	Thallium	81	203	202,972329	41,892	0,2925	204,3833
Tl	Thallium	81	205	204,974412	100	0,7048	204,3833
Pb	Blei	82	204	203,973029	2,7	0,014	207,2
Pb	Blei	82	206	205,974449	46	0,241	207,2
Pb	Blei	82	207	206,975881	42,2	0,221	207,2
Pb	Blei	82	208	207,976636	100	0,524	207,2
Bi	Wismut	83	209	208,980383	100	1	208,980383
Th	Thorium*	90	232	232,03805	100	1	232,03805
U	Uran*	92	234	234,040946	0,0055	0,000054	238,0289
U	Uran*	92	235	235,043923	0,73	0,007204	238,0289
U	Uran*	92	238	238,050783	100	0,992742	238,0289

B Peak-Anzahl simuliertes fragmentiertes Spektrum

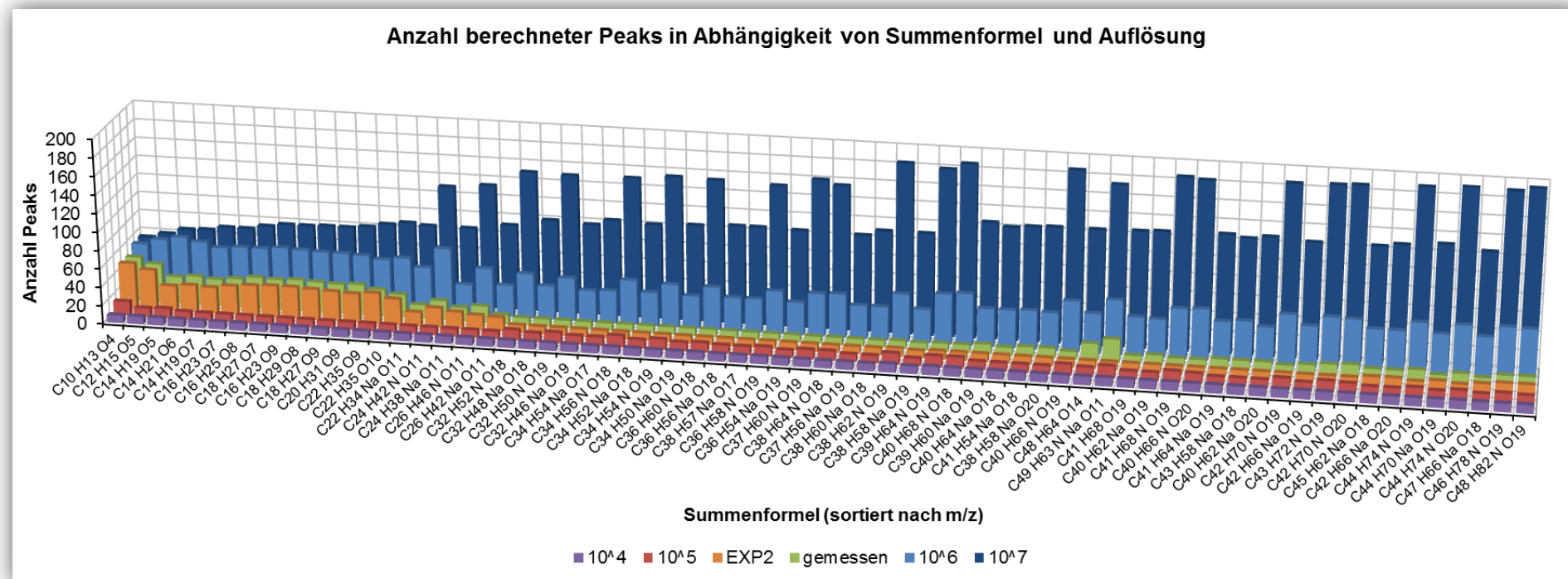


Abbildung A-1: Anzahl berechneter Peaks pro Summenformel mit der multinomialen Methode des fragmentierten Spektrums

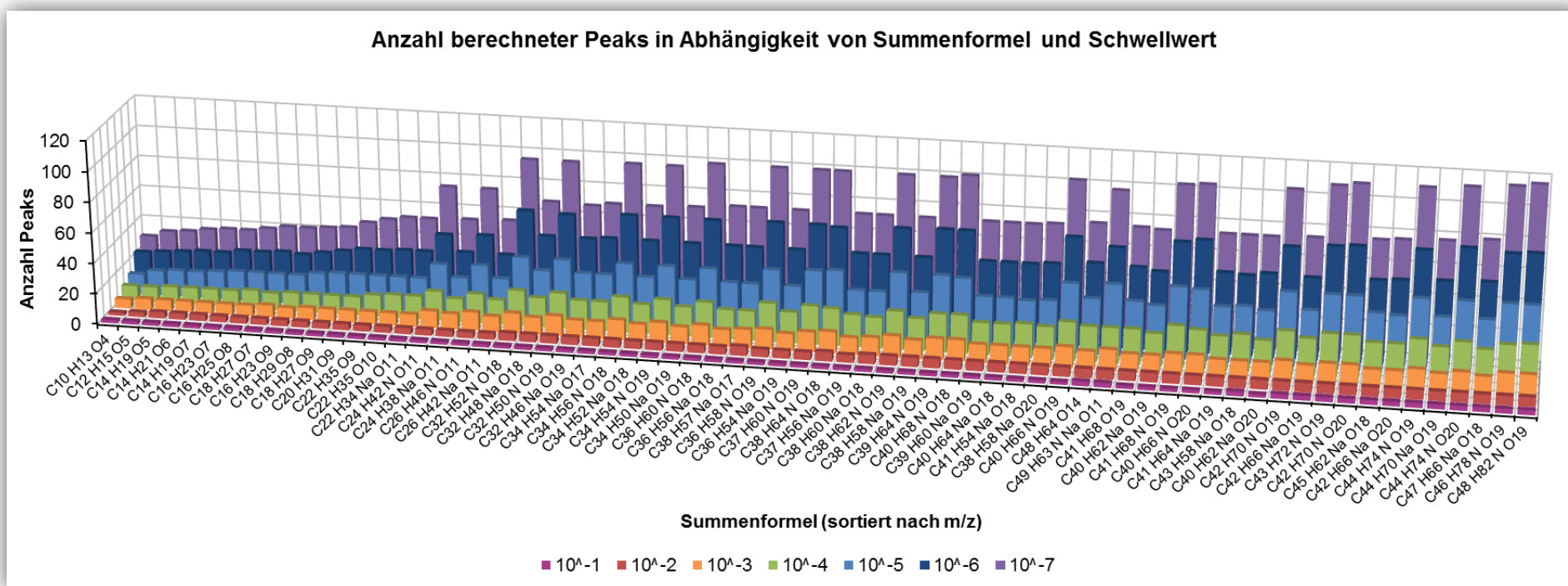


Abbildung A-2: Anzahl berechneter Peaks pro Summenformel mit der 2D-FT Methode des fragmentierten Spektrums

C Laufzeiten Multinomial und 2D-FT Methode

Tabelle A-2: Laufzeiten der multinomialen und 2D-FT Methode für verschiedene Summenformeln

Es sind die Laufzeiten in Millisekunden für die fünf Summenformeltypen aufgelistet. MN steht für multinomial und FT für den 2D-FT Algorithmus. Es wurden für die multinomiale Methode ein Schwellwert von 0,001 und für die 2D-FT Methode eine Auflösung von 10^7 verwendet. Typ 1: $C_n H_{2n+2}$, Typ 2: $C_n H_{2n+2+m} O_n$, Typ 3: $C_n H_{2n+2+m} O_n N_m$, Typ 4: $C_n H_{2n+2+m} O_n N_m S_m$, Typ 5: $C_n H_{2n+2+m} O_n N_m S_m F_{e_m/2}$.

(n,m)	Typ 1		Typ2		Typ 3		Typ4		Typ5	
	MN	FT	MN	FT	MN	FT	MN	FT	MN	FT
(10, 2)	10	1193	1	867	1	720	1	2683	1	3899
(20, 4)	6	264	1	263	1	719	1	3202	1	14209
(30, 6)	3	120	1	271	5	287	1	2573	32	3904
(40, 8)	3	123	5	295	5	315	10	2556	93	4837
(50, 10)	3	126	10	166	5	297	11	2880	114	5818
(60, 12)	6	133	32	140	43	140	36	2780	151	7137
(70, 14)	3	126	135	138	130	150	1	2854	504	1592
(80, 16)	3	156	489	137	483	252	509	809	1045	1521
(90, 18)	10	74	1726	258	1737	270	1778	823	3070	1522
(100, 20)	3	83	5773	139	5783	258	5851	8	7302	1570
(110, 22)	6	80	18309	136	18424	266	18409	890	21399	1531

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, den 28.02.2014

Bianca Liebscher